European Journal of Operational Research 265 (2018) 178–194

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Are objectives hierarchy related biases observed in practice? A meta-analysis of environmental and energy applications of Multi-Criteria Decision Analysis

Mika Marttunen[a,*], Valerie Belton[b], Judit Lienert[c]

[a] Swiss Federal Institute of Aquatic Science and Technology, Eawag, P.O. Box 611, CH-8600 Duebendorf, Switzerland
[b] Business School, University of Strathclyde, 199 Cathedral Street, Glasgow G4OQU, Scotland, UK
[c] Swiss Federal Institute of Aquatic Science and Technology, Eawag, P.O. Box 611, CH-8600 Duebendorf, Switzerland

ARTICLE INFO

Article history:
Received 7 May 2016
Accepted 13 February 2017
Available online 7 March 2017

Keywords:
Behavioural OR
Decision analysis
Decision processes
Multiple criteria analysis
OR in environment and climate change

ABSTRACT

Procedural and behavioural biases have received little attention in recent Multi-Criteria Decision Analysis (MCDA) research. Our literature review shows that most research on biases was done 15–30 years ago. This study focuses on biases that are introduced at an early stage of MCDA when building objectives hierarchies and their effect on the weights. The main objective is to investigate whether prior findings regarding such biases, which were mostly based on laboratory experiments, can be found in real-world applications. We conducted a meta-analysis of the objectives hierarchies and weight elicitation procedures in 61 environmental and energy MCDA cases. Relationships between the structural characteristics of the objectives hierarchy and assigned objectives' weights were analysed with statistical tests. Our main research questions were: (i) How does hierarchy size and structure affect the objectives' weights? (ii) How are weights distributed across economic, social and environmental objectives? (iii) Is there support for the equalising bias? Our findings are mostly aligned with earlier research and suggest that the hierarchy structure and content can substantially influence weight distributions. For example, hierarchical weighting seems to be sensitive to the asymmetry bias, which can occur when a hierarchy has branches that differ in the number of sub-objectives. We found no evidence for the equalising bias. We highlight issues deserving more attention when developing objectives hierarchies and eliciting weights. The research demonstrates the potential to use meta-analysis, which has not previously been used in this way in the MCDA field, to learn from a collection of applications.

1. Introduction

Different types of cognitive and behavioural biases play an important role in both unaided and aided decision making. Multi-Criteria Decision Analysis (MCDA), which aims to help people to make decisions that are in agreement with their values and understanding of the problem (Keeney & Raiffa, 1976), can be subject to several behavioural and procedural biases (Montibeller & von Winterfeldt, 2015). These biases can occur in all phases of the decision making process (problem structuring, impact assessment, preference modelling and drawing conclusions and at worst lead to incorrect recommendations.

Behavioural and procedural biases have received surprisingly little attention in the practice of MCDA research in recent years (Hämäläinen, 2015; Hämäläinen & Alaja, 2008; Montibeller & von Winterfeldt, 2015; Pöyhönen & Hämäläinen, 2000). The literature review informing this study shows that much of this research was carried out 15 to 30 years ago but there are strong indications of renewed attention in operational research (OR) (Franco & Hämäläinen, 2016). In addition to recent reviews (Hämäläinen, 2015; Montibeller & von Winterfeldt, 2015), a special issue of EJOR focused on behavioural aspects of OR (Volume 249, issue 3).

Most of the research about procedural and behavioural biases has been conducted with students in hypothetical decision situations. Laboratory experiments differ in many respects from real situations and are prone to several types of errors (Pöyhönen, 1998). We focus on the biases which are associated with objectives hierarchies (also called decision hierarchies or value trees). The main objective of this study is to investigate whether the findings of

* Corresponding author. Present address: Finnish Environment Institute, SYKE, PB 140, 00251 Helsinki, Finland.

E-mail addresses: mika.marttunen@eawag.ch, mika.marttunen@environment.fi (M. Marttunen), val.belton@strath.ac.uk (V. Belton), judit.lienert@eawag.ch (J. Lienert).

http://dx.doi.org/10.1016/j.ejor.2017.02.038
0377-2217/

prior studies are replicated in real-world environmental and energy MCDA applications.

Our main research questions are: (i) How do hierarchy size and structure affect the objectives' weights? (ii) How are weights distributed across economic, social and environmental objectives? and (iii) Do people have a tendency to give equal weights to the objectives? We applied a meta-analysis approach, which is widely used in the medical, social and ecological sciences (Arnqvist & Wooster, 1995; Koricheva, Gurevitch, & Mengersen, 2013; Petitti, 1994), but to our knowledge has not yet been applied in MCDA. Therefore, an additional objective is to examine the benefits and pitfalls of meta-analysis in the setting of this study.

The study has three main phases; first, earlier research related to behavioural and procedural biases in MCDA was reviewed; second, a literature search was carried out and 61 cases were selected for the further analysis; and thirdly statistical tests were conducted to analyse relationships between the structure and size of the hierarchy and the weights assigned to objectives.

The paper is organised as follows. Section 2 presents the basics of MCDA methods, defines the terminology and reviews the literature on behavioural and procedural biases. In Section 3 we introduce three research questions and seven related analyses. Section 4 describes our research methods and how meta-analysis was realised in this study. Section 5 presents the results, starting with general information about the selected cases and the objectives hierarchies used, followed by the analyses of the research questions. In Section 6 we discuss the practical relevance of the results and present recommendations on how to diminish the risk of biases. We also evaluate the pros and cons of meta-analysis in the MCDA context and make suggestions for follow-up studies. Section 7 concludes the article.

2. Literature review

2.1. Multi-Criteria Decision Analysis (MCDA)

The use of MCDA in environmental applications has increased considerably in number and diversity during the last decade (Huang, Keisler, & Linkov, 2011a; Keisler & Linkov, 2014). Several reviews have been published, focusing on: motivations for the applications; the nature of the problems addressed; methods used; and the appropriateness of the MCDA approach (e.g. Ananda & Herath, 2009, Balasubramaniam & Voulvoulis, 2005, Hajkowicz & Collins, 2007, Huang et al., 2011a, Mendoza & Martins, 2006).

There are many methods and associated softwares in the field of MCDA today. The methods differ in terms of underlying assumptions and principles, and apply different procedures for scoring, weighting and aggregation (Belton & Stewart, 2002). In many recent MCDA applications, the major aim is not simply to make a choice between the alternatives but to use the systematic MCDA framework to explore objectives and alternatives, facilitate communication, enhance social learning and support consensus finding (e.g. Antunes, Karadzic, Santos, Beca, & Osann, 2011; Bana E Costa, Da Silva, & Correia, 2004; Marttunen, Mustajoki, Dufva, & Karjalainen, 2015).

This study pays particular attention to the central elements of any applied MCDA, the structure of the objectives hierarchy and the weights assigned to the specified objectives. Objectives hierarchies define the variety of concerns and the aims that decision makers wish to achieve. The hierarchy is the basis for the evaluation, guiding the search for information, influencing the comparison of alternatives and how preferences are elicited (Borcherding & von Winterfeldt, 1988).

The relative importance of the objectives is a key concept in MCDA and is usually captured by assigning weights to the objectives. The interpretation of the weights differs according to the method. In MAVT (Multi-Attribute Value Theory, Keeney & Raiffa, 1976) weights are scaling factors which determine the relative added value associated with the impact range defined for each criterion for an individual decision maker (e.g. Eisenführ, Weber, & Langer, 2010). The impact range is the difference between the best and worst alternative (local scale) or best and worst possible outcome (global scale) with respect to each attribute. The performance of alternative with respect to each objective is transformed to the defined scale (usually 0–1 or 0–100); these scores are then aggregated using the specified weights to give an overall performance value for each alternative. In Analytic Hierarchy Process (AHP, Saaty, 1980), the interpretation of the weights is less clear (Belton, 1986; Pöyhönen & Hämäläinen, 2001; Salo & Hämäläinen, 1997). Outranking methods, such as ELECTRE (Roy, 1991) and PROMETHEE (Brans, Vincke, & Mareschal, 1986), focus on pairwise comparisons of alternatives and on outranking relations. PROMETHEE does not provide any guidelines to determine the weights and, for instance, AHP has been combined with PROMETHEE (Macharis, Springael, De Brucker, & Verbeke, 2004).

MCDA methods differ greatly with regards to how clearly the impact ranges are presented in weight elicitation. If the ranges are not taken into account, then the weights can represent general values or attitudes toward objectives rather than trade-offs between them (Fischer, 1995). Swing (von Winterfeldt & Edwards, 1986) and trade-off (Keeney & Raiffa, 1976) methods explicitly present the ranges in the weight elicitation procedure. In SMART (von Winterfeldt & Edwards, 1986), ratio (Edwards, 1977) and point allocation methods the decision maker directly assigns weights to objectives. These methods do not explicitly incorporate ranges when weight judgments are derived (von Nitzsch & Weber, 1993; Weber and Borcherding, 1993). In AHP pairwise comparisons typically do not include the impact ranges.

Authors use different terms for the elements of the objectives hierarchy. For instance, an objective is often called a criterion or an attribute. In this article, an attribute is a variable which measures, or informs the measurement of, the alternatives impacts on a specific objective. Fig. 1 illustrates the terms we use in association with an objectives hierarchy. In addition, we use the term parental objective to refer to any objective which has sub-objectives. The term participant is used to describe persons engaged with the MCDA process. They can be, for instance, a decision maker, a representative of the stakeholder group or an expert. A weight profile means a set of weights given to the objectives, either by one participant or determined by aggregating weights over a group of people.

Weights can be elicited either hierarchically or non-hierarchically (Fig. 2). Hierarchical weighting can be realised either top-down or bottom-up. The latter is more recommendable because participants' understanding of alternativeśimpact ranges can be better in the bottom-up than in the top-down approach. In the non-hierarchical procedure, weights are first assigned across all the lowest-level objectives. The weight of an objective at a higher level of the hierarchy is a sum of the weights assigned to the sub-objectives belonging to that branch, if additive aggregation is used.

2.2. Review of empirical research

Several earlier studies have shed light on how features of the objectives hierarchy and weight elicitation procedure can affect participants' judgments (Table S1 in the supplementary material, later used abbreviation SM). Most of these studies were controlled experiments conducted with students. In the Sections 2.2.1 and 2.2.2 we summarise the main results of the studies, which were key in defining our research questions and in the interpretation of associated analyses.

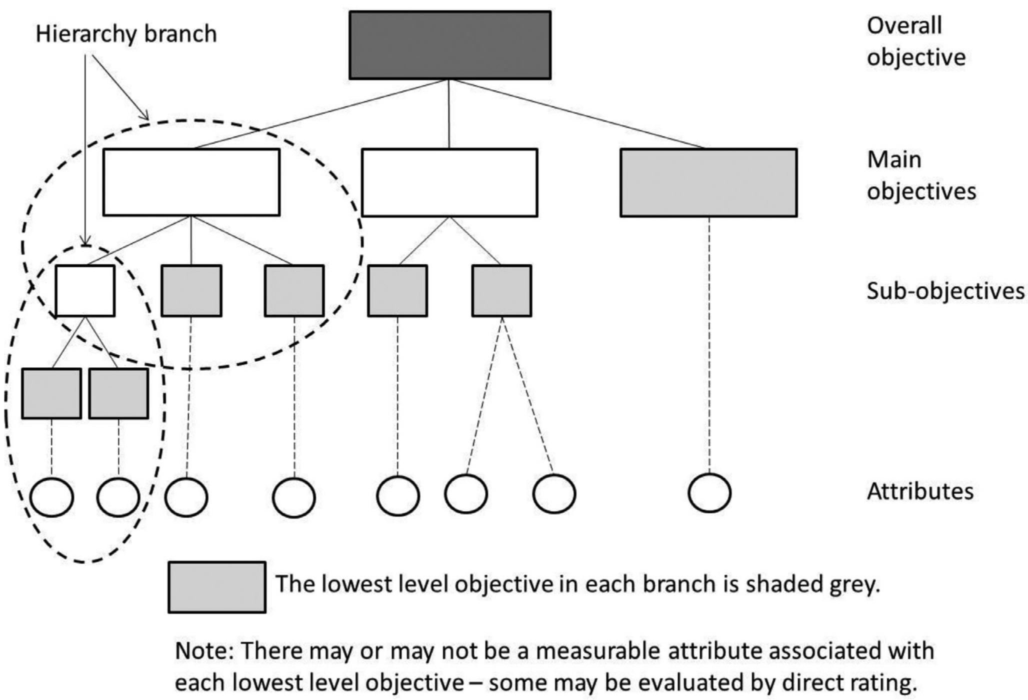


Fig. 1. Terminology related to an objectives hierarchy.

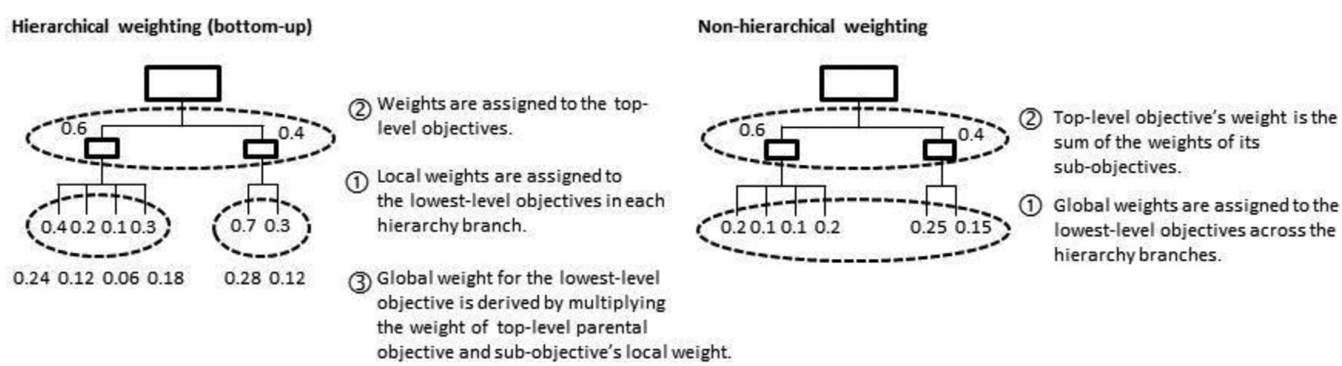


Fig. 2. Hierarchical (left) and non-hierarchical weighting procedures (right) with a numerical example.

2.2.1. Biases related to the objectives hierarchy

The splitting bias refers to the phenomenon where dividing an objective into two (or more) objectives in a branch of a hierarchy produces an increase in the overall weight of that branch when non-hierarchical weighting is used (Hämäläinen & Alaja, 2008). It is perhaps the most studied bias in the MCDA literature; five of fifteen identified studies focused on it (Table S1 in the SM). All splitting bias experiments reported differences between the split and unsplit objectives' weights (Borcherding & von Winterfeldt, 1988; Hämäläinen & Alaja, 2008; Pöyhönen & Hämäläinen, 1998; Pöyhönen, Vrolijk, & Hämäläinen, 2001; Weber, Eisenführ, & Von Winterfeldt, 1988). Avoiding the splitting bias is difficult; it also has been detected in situations where participants have been instructed to avoid it (Hämäläinen & Alaja, 2008; Pöyhönen & Hämäläinen, 2000).

Although the evidence of the splitting bias is based on studies with non-hierarchical weighting procedures, it can also occur in hierarchical weighting (Pöyhönen et al., 2001). People seem to avoid giving extremely low or extremely high weights to main objectives and, as a result, a main objective with fewer sub-objectives will have to split its weight less, resulting in higher per-objective weight than branches that have more sub-objectives (Hobbs & Meier 2000, p. 77). Schuwirth, Reichert, and Lienert (2012, also, Scholten, Schuwirth, Reichert, & Lienert, 2015, Zheng, Egger, & Lienert, 2016) tried to mitigate against the overweighting of hierarchy branches presented in more detail by not showing the sub-objectives before weights were assigned to the main objectives.

Number of objectives: The distribution of weights is influenced by the number of objectives due to the normalisation of weights so that they sum to one. Salo and Hämäläinen (1997) show how the maximum weight of objectives depends on the upper and lower bounds of the used scale in AHP. For instance, when Saaty's original rating scale 1–9 was used, the theoretical maximum weights are 0.75 and 0.50 for four and ten objectives respectively. A similar phenomenon occurs, although weaker (i.e. maximum weights are higher), in weighting methods where direct rating on a 0–100 scale is used, if a weight is allocated to all objectives. Weber et al. (1988) found that the ratio between the largest and smallest weight increases considerably as the number of objectives increases and the same phenomenon was demonstrated mathematically by Pöyhönen and Hämäläinen (1998). The weight assigned to a single objective averaged over a large group of participants tends to follow the rule 1/n, where n is the number of objectives

(Fox & Clemen, 2005; Jacobi & Hobbs, 2007; Pöyhönen & Hämäläinen, 1998; Pöyhönen & Hämäläinen, 2001). This rule resembles the equalising bias which, however, describes the individuals' behaviour instead of means over large groups (see Section 2.2.2).

Location of the objective: Experiments show that an objective receives more weight, using hierarchical weighting, if it is presented higher in the hierarchy[1] (Borcherding & von Winterfeldt, 1988; Jacobi & Hobbs, 2007; Pöyhönen & Hämäläinen, 2000). The phenomenon has been explained by the "anchor-and-adjust" heuristic in which the participants start with an equal allocation of weights to each objective as anchors and then adjust the weights insufficiently (Jacobi & Hobbs, 2007). Another explanation for the higher weight is that there are fewer objectives at the higher levels of the objectives hierarchy (Pöyhönen & Hämäläinen, 2000).

Type and label of objective: There is some evidence that environmental and social objectives receive higher weights than economic objectives (Gregory et al., 2012; Keeney, 2002). A possible reason is that these objectives are generally considered morally or ethically more important or socially more acceptable (Gregory et al., 2012; Stillwell, Von Winterfeldt, & John, 1987). In particular, this can be the case if the consequences are not clearly presented (Keeney, 2002). Stillwell et al. (1987) remark that the labels of main objectives, which can be abstract, may not be psychologically equivalent to the set of more concrete lowest-level objectives. Further evidence about the influence of the label of the objective is provided by Hämäläinen and Alaja (2008). This is discussed in more detail in section 6.2.

Type of attribute: Attributes are used to measure how well alternatives meet the objectives. Keeney and Gregory (2005) divide them into three groups: natural, constructed and proxy attributes. A proxy attribute is an indirect measure (or indicator) to assess the degree to which an objective is achieved. The challenge in the use of the proxies is that determining the relation between the levels of the proxy and the fundamental objective is left to the participants who may not have enough expertise for that. Overall, understanding this relationship can be cognitively very demanding and therefore lead to the use of simple heuristics (Keeney & Raiffa, 1976). For instance, in an experimental study of preferences for pollution control alternatives, participants systematically overweighted the proxy attribute, emissions level, compared to the fundamental objective, illness level (Fischer, Damodaran, Laskey, & Lincoln, 1987). Proxy attributes are often used in environmental applications due to complexity of the systems and associated lack of knowledge (e.g. Keeney, 2007, p. 119). For such difficult environmental valuations it has been suggested to use expert assessments at the more technical lower hierarchy levels and ask for preferences from stakeholders or the population only for the higher levels of the objectives hierarchy, which represent the major societal trade-offs (Reichert, Langhans, Lienert, & Schuwirth, 2015). However, this procedure may introduce additional biases as discussed in section 2.2.2 as lay people can then have more difficulties understanding the full meaning of the higher level objectives.

2.2.2. Biases related to weighting procedures

The range insensitivity bias refers to the phenomenon that participants do not sufficiently adjust their weights if the range of attributes is changed (von Nitzsch & Weber, 1993). For instance, if the difference between the lowest and highest costs increases from 100,000 € to 500,000 € after the addition of a new alternative and all other ranges remain the same, then the weight of the cost objective should increase. However, experiments suggest that this is not always the case. This experimental finding indicates that participants may generally not sufficiently consider the range when giving their weight estimates. The range insensitivity bias can be reduced by explicitly presenting the impact ranges (Fischer, 1995; von Nitzsch & Weber, 1993) and by educating participants (Pöyhönen & Hämäläinen, 2000). It is notable that this imperfect adjustment of the objectives' weights can explain the splitting bias (Weber et al., 1988).

The equalising bias suggests that people tend to give equal weights to all objectives (Montibeller & von Winterfeldt, 2015). The equalising bias is less studied in MCDA and evidence comes mainly from resource allocation and probability assessment experiments in which people tend to anchor on a uniform distribution of resources or probability (e.g. Bardolet, Fox, & Lovallo, 2011, Benartzi & Thaler, 2001, Fox & Clemen, 2005). Huberman and Jiang (2006) noticed that the tendency to allocate resources equally decreased when participants had higher expertise. It is not clear if these results can be transferred as such to objectives' weights. However, Jacobi and Hobbs (2007) took the equalising bias as a starting point in their model for estimating and correcting objectives hierarchy induced biases. Their results supported the "anchor-and-adjust" heuristic in weight elicitation. There is also evidence which is not consistent with the equalising bias. For instance, in a splitting bias experiment Weber et al. (1988) found that the weights were not assigned evenly across the objectives. Furthermore, the ratio between the highest and lowest weight increased substantially as the number of objectives increased. Additionally, there is also evidence that in personally important decisions (e.g. in the choice of a house) people have a few highly weighted objectives and a number of low-weighted objectives which receive less attention (Fasolo, Mcclelland, & Todd, 2007; Saad & Russo, 1996).

Hierarchical and non-hierarchical weighting: Weights of objectives can be elicited either using hierarchical or non-hierarchical weighting (see explanation in section 2.1, Fig. 2). Most often the weighting is realised hierarchically because the number of simultaneous comparisons is lower than in non-hierarchical weighting (Pöyhönen et al., 2001). In all experiments which have compared these weighting procedures, hierarchically generated weights had a higher variance than non-hierarchically generated weights (Jacobi & Hobbs, 2007; Sayeki & Vesper, 1971; Stillwell et al., 1987).

Weighting method: A number of studies compared the convergence of different weighting methods (Belton, 1986; Bottomley & Doyle, 2001; Cook & Stewart, 1975; Doyle, Green, & Bottomley, 1997; Fischer, 1995; Pöyhönen & Hämäläinen, 2001; Schoemaker & Waid, 1982). The results regarding the convergence of swing, direct rating and trade-off method vary in different studies. For instance, Fischer (1995) found that the trade-off method yielded higher weights for the most important objective than swing and direct rating, whereas Pöyhönen and Hämäläinen (2001) concluded that the weights of direct rating, swing and trade-off did not differ from each other. In AHP, weight distribution has been found to be larger than in direct rating, swing and trade-off (Pöyhönen & Hämäläinen, 2001; Schoemaker & Waid, 1982), and weights were more unevenly dispersed than in SMART (Belton, 1986). The experiments show that the numerical scale which is used in AHP (e.g. 1–9 scale, balanced scale) affects the weight ratios (Salo & Hämäläinen, 1997) and can explain some differences between methods (Lienert, Duygan, & Zheng, 2016). These findings are partly in line with the proposition of van Ittersum, Pennings, Wansink, and Van Trijp (2007) who divided the weighting methods into three groups

[1] For example, if objectives X and Y are presented in the top level of a hierarchy, the sum of their weights would be greater than that allocated to Z, having X and Y as sub-objectives.

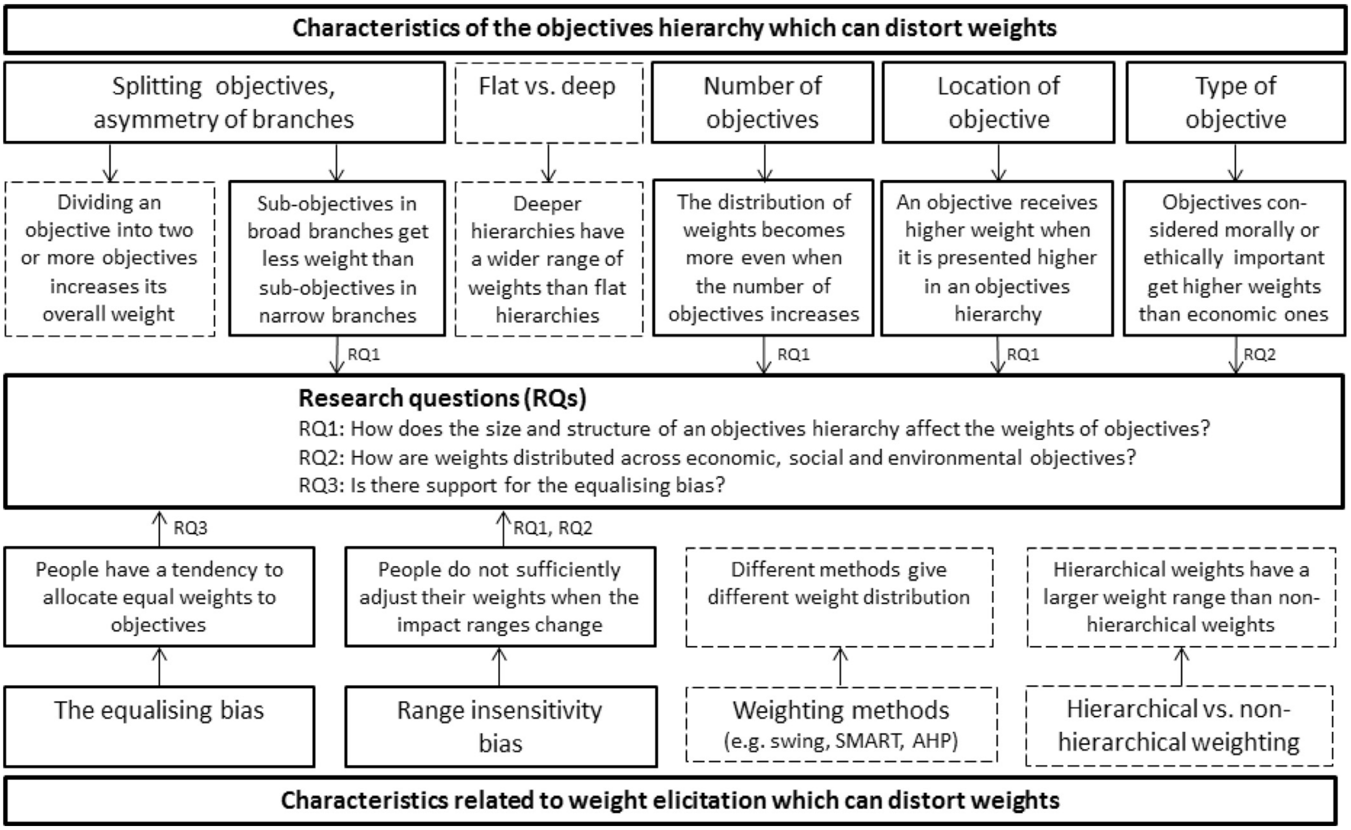


Fig. 3. Behavioural and procedural biases related to the size and structure of the objectives hierarchy and to weight elicitation. Biases marked with dashed line were not studied.

according to their salience, relevance and determinance[2], and suggested that methods belonging to the same group give more identical weights than methods in different groups.

2.2.3. Limitations of experiments

Laboratory experiments differ in many respects from real situations and are prone to several types of errors (Pöyhönen, 1998). Most reported experiments used students whose task was to give preferences in hypothetical decisions with only few objectives (Table S1 in the SM). Typically, students are better-educated and numerically better trained than the general public. Particularly, students with a background in economics or applied mathematics, which were the participants in many earlier studies, may feel much more comfortable with numbers than people in general. It is also possible that students have a weak motivation for thinking carefully about their preferences (Pöyhönen & Hämäläinen, 2000).

It is not unproblematic to use lay people in experiments either, because the methods can be too difficult and the cognitive load can become too high (Cook & Stewart, 1975; Hämäläinen & Alaja, 2008). It is also possible that in the experiments people become more aware of the studied phenomenon (Fischer, 1995), which may reduce the generalisability of the results. For instance, Hämäläinen and Alaja (2008) observed that some students took the splitting bias experiment as a calculation exercise aiming to minimise the bias. For these reasons, it is important to investigate whether the biases demonstrated in laboratory conditions can also be found in real-world cases. If the experimental results are valid, similar phenomena should be found in the cases analysed in this study.

All behavioural research in MCDA faces the same challenge. If two methods give different results we cannot know which weights best reflect the participants' true opinions. It is therefore impossible to say which biases are "good" and which are "bad" (Pöyhönen & Hämäläinen, 2000). Similarly, it is difficult to find out whether the assigned weights to objectives really reflect participant's true opinions.

3. Research questions

We set up three research questions and seven specific analyses. Four analyses aimed to find out how the characteristics of the objectives hierarchy affect the weights (1a–1d), two focused on the weight distribution of environmental, economic and social objectives (2a–2b) and one on the equalising bias (3). The research questions and analyses were developed based on previous research (Fig. 3).

- Research question 1: How does the size and structure of the hierarchy affect the objectives' weights?
  - 1a. Total number of lowest-level objectives and their highest weight.
  - 1b. Total number of lowest-level objectives and number of objectives getting very low weights.
  - 1c. Mean global weights of the sub-objectives in the largest and smallest branch.
  - 1d. Location of the most important objective in the objectives hierarchy.
- Research question 2: How are weights distributed across economic, social and environmental objectives?
  - 2a. Weight ratios of environmental and economic objectives.

[2] Salience reflects the degree of ease with which objectives come to mind or are recognized when thinking about or seeing a certain object. The relevance of objectives is largely determined by personal values and desires and reflects the importance of objectives for individuals. The determinance of an objective reflects the importance of an objective in judgment and choice (the difference in the objectives' impacts is considered). The weight elicitation methods belonging to the relevance group include e.g. direct rating, point allocation and AHP, and to the determinance group e.g. trade-off and swing.

Table 1
Data collected from the selected papers.

Type of information	Characteristics
General	Authors; Year; Country of application; Country of first author; Application area.
Structure of the objectives hierarchy	Number of main objectives; Number of hierarchy levels; Number of hierarchy branches at the top level of the hierarchy; Number of lowest-level objectives.
Number of objectives	Economic, Technical, Socio-economic, Social, Environmental, Risks, Other objectives.
Costs in the hierarchy	Main objective (either divided or not divided into sub-objectives); Sub-objectives of the economic objective; Not included in the analysis.
MCDA method	Name of the method (e.g. MAVT, AHP, ELECTRE, PROMETHEE).
Weight elicitation technique	Name of the technique (e.g. ratio, SMART, swing, trade-off).
Weight elicitation procedure	Bottom-up; Top-down; Hierarchical weighting; Non-hierarchical weighting; Unclear.
Source of preferences	Decision makers; Policy makers; Experts; Students; Hypothetical; Authors; Unclear.
Method to collect preferences	Questionnaire or Survey; Workshop; Interviews; Literature; Unclear.
Presentation of objectives' weights	Individually; Group mean; Mean across all participants; Number of weight profiles.

- 2b. Weight ratios of social and economic objectives
- Research question 3: Is there support for the equalising bias?
  - 3. Ratio of the lowest and highest weights of main objectives.

4. Research methods

4.1. Literature search

We started the search using "multi-criteria" and "environment". These searches gave us more than 2000 hits in Web of Science. Therefore, we used following key words to narrow the search (MCDA, MAVT, MAUT, AHP) AND (environment, water, forestry, fishery, energy) AND (case, application). The potential articles were quickly scanned. We excluded the cases which did not present the weights of objectives as well as the papers which used very simple hypothetical weight profiles, such as balanced weights. To get a sample which covered a large variety of hierarchy sizes, different methods to gather preference information (interviews, questionnaires) and different weight elicitation techniques (e.g. swing, AHP) we conducted several Google Scholar searches using the search words limited to the period of 2013–2015. Higher priority was given to cases which engaged several stakeholders, because one article that presents ten weight profiles from ten different people provides as much information about the objectives' weights as ten articles that present only one weight profile. The selected cases included six Finnish and eight Swiss cases which we were familiar with and for which data was easily available.

Finally, 59 papers and 61 objectives hierarchies (later called cases) from these papers were selected for the analysis (see the SM for the references of all cases). From each paper we collected information about the features of the hierarchy and weight elicitation procedure (Table 1, see Table S-1 in the SM). We discuss the limitations of the search procedure in Section 6.5.

4.2. Meta-analysis and statistical analyses

Meta-analysis is the application of statistical procedures to collections of empirical findings from individual studies for the purpose of integrating, synthesising and making sense of them (Wolf, 1986). It has been largely used, for instance, in the social and medical sciences and in ecology and economics. The main phases of meta-analysis are (i) a comprehensive review of the literature, (ii) systematic analysis of the quality and content of each study, and (iii) analyses of the combination of data or results from cases studies and the drawing of appropriate conclusions. Matarazzo and Nijkamp (1997) provide an overview of different types of meta-analyses in environmental case studies. Meta-analysis has increasingly been used to synthesise the results of environmental valuation studies (Lindhjem & Navrud, 2008; Nelson & Kennedy, 2009) but to our knowledge the only application to date in the MCDA field is a qualitative study to evaluate the legitimacy and quality of five MCDA processes in Norway (Wenstøp and Seip, 2001).

This research seeks to demonstrate the potential to learn about aspects of MCDA in practice from a larger scale, quantitative meta-analysis. Nevertheless, we recognise that the extent of our study is modest compared with sophisticated meta-analyses in the fields of psychology, medicine and economy, and that it differs in three main ways:

(i) Aim of the study: We used data from real-world applications to investigate whether the same biases that were observed in prior laboratory experiments are replicated in the real-world. This is a less ambitious aim than in those cases where meta-analysis is conducted to produce, for example, a more precise estimate of a medical treatment or to develop an econometric model based on the results of multiple studies.

(ii) Selection of the cases: Careful selection of the cases which are included in a study is an essential phase of meta-analyses. If the meta-analysis is used to integrate results of different studies, it is important to evaluate the quality of the included studies, and often poor quality of studies are excluded (Meline, 2006). In this study, the selected cases used different approaches to weight elicitation (see supplementary material, Table S-2) but there was nothing which indicated that these were poorly or inappropriately applied.

(iii) Applied statistical analyses: Due to the different aims of this study and the nature of the material, we used much simpler statistical analyses (e.g. Spearman's Rank Correlation, Related-Samples Sign Test) than more standard meta-analyses (see e.g. Hedges & Olkin, 1985).

For the statistical analysis we used SPSS version 20. In three analyses we used Spearman's Rank Correlation and in three the Related Samples Sign test (Table 2). It was not appropriate to include all 61 cases in all analyses. The suitability of the cases for each analysis was defined by the content and structure of the hierarchy as well as the number of weight profiles. Note that both the participants' individual weights and the mean weights of a group of people were used in the analyses 1a,b and 2a,b.

5. Results

5.1. Characteristics of the cases

5.1.1. Country and application area

The cases came from 25 countries (Fig. 4). The highest numbers came from the USA, Switzerland and Finland. Half of the papers were from the period 2011–2015 (Fig. 4). This mainly reflects the strong recent increase in the number of environmental MCDA applications (Huang et al., 2011a), but is partly also a consequence

Table 2
Number of cases (total $n = 61$) and weight profiles (total $n = 230$) and statistical methods used in the meta-analysis.

	Description of analyses	Description of data	Statistical test	Number of cases / weight profiles included in the test; criteria for inclusion
1a	Total number of lowest-level objectives and weight of the most important objective.	Weight data is not normally distributed. There is a monotonic decreasing relationship between the number of lowest-level objectives and the highest weight.	Spearman's Rank Correlation	61/230 All cases and weight profiles were used.
1b	Total number of lowest-level objectives and number of objectives getting very low weights.	Proportion of low weights is not normally distributed. Monotonic increasing relationship between the number of lowest-level objectives and the number of objectives having weights ≤ 0.05.	Spearman's Rank Correlation	61/230 All cases and weight profiles were used.
1c	Mean global weights of the sub-objectives in the largest and smallest branch.	Weight data is not normally distributed. Global weights in the smallest and largest branch are not independent because of the normalisation (sum of weights = 1).	Related-Samples Sign Test	23/103 Cases with asymmetric objectives hierarchies and which present several weight profiles.
1d	Location of the most important objective in the objectives hierarchy	Statistical test not conducted.	None	16/65 Cases with asymmetric objectives hierarchies and which present several weight profiles. Cases having flat branches were excluded to improve comparability.
2a	Weight ratio of environmental and economic objectives.	Economic weights are not normally distributed and their distribution is not symmetric with environmental weight distributions. Weights are not independent because of the normalisation (sum of weights = 1).	Related-Samples Sign test	31/124 Cases which include both environmental and economic main objectives.
2b	Weight ratio of social and economic objectives.	As 2a.	Related-Samples Sign test	25/96 Cases which include both both social and economic main objectives.
3	Ratio of the lowest and highest weights of main objectives (the equalising bias).	Lowest and highest weights are not normally distributed and distributions are not symmetric. Weights are not independent because of the normalisation (sum of weights = 1).	Spearman's Rank Correlation	10/96 Cases which have two to four main objectives related to e.g. environmental, social, socio-economic or economic objectives. Several weight profiles are presented.

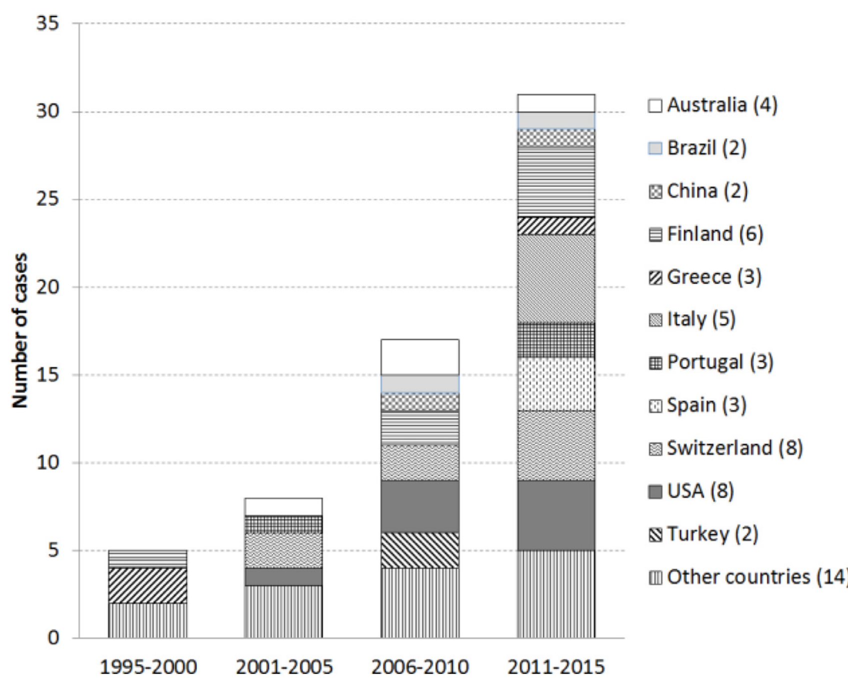


Fig. 4. Publication year and country of the cases analysed in this study. The countries which have less than two articles are not listed separately. Note that 1995–2000 is a six years period.

of the search and selection procedure (see Section 4.1). The most common application areas were water resources management (22 cases), energy planning (15), wastewater management (4) and forest or wetland management (3).

5.1.2. Objectives hierarchies

The selected cases widely varied in their hierarchy size and structure (Table 3). The number of lowest-level objectives ranged from 3 to 51, the mean being 14.6 (median 13). The deepest hierarchy consisted of five hierarchy levels (the overall objective was not included in the count of hierarchy levels) and the flattest hierarchies had only one level. Environmental objectives were included in 56, economic in 51 and social objectives in 39 cases. The number of environmental objectives is somewhat higher than the number of economic and social objectives, means being 5.7, 3.1 and 3.8, respectively.

Table 3
Characteristics of the objectives hierarchies in the analysed cases ($n = 61$). The distinction between economic, socio-economic, social, environmental and technical objectives is based on the terms used in the papers and our judgments.

	Mean	Standard deviation	Minimum	Maximum
Number of top level objectives	4.7	3.1	2	18
Number of lowest-level objectives	14.6	8.2	3	51
Total number of objectives	19.3	11.4	4	73
Number of hierarchy levels	2.3	0.9	1	5
Number of lowest-level economic objectives ($n = 51$)	3.1	2.4	0	10
Number of lowest-level socio-economic objectives ($n = 10$)	3.7	4.6	0	11
Number of lowest-level social objectives ($n = 39$)	3.8	2.8	0	17
Number of lowest-level environmental objectives ($n = 56$)	5.7	7.3	1	51
Number of lowest-level technical objectives ($n = 14$)	3.7	3.0	0	12
Number of weight profiles presented in the paper	3.8	4.4	1	25

Table 4
Typology of the objectives hierarchies and number of cases in each hierarchy type.

Breadth/Depth	Narrow ≤ 5 LLOs[1]	Medium 6–10 LLOs	Broad 11–15 LLOs	Very broad > 15 LLOs	Total
Flat (1 level)	2	5	3	1	11
Medium (2 levels)	1	12	9	6	28
Deep (≥ 3 levels)	0	2	8	12	22
Total	3	19	20	19	61

[1] LLO is the lowest-level objective

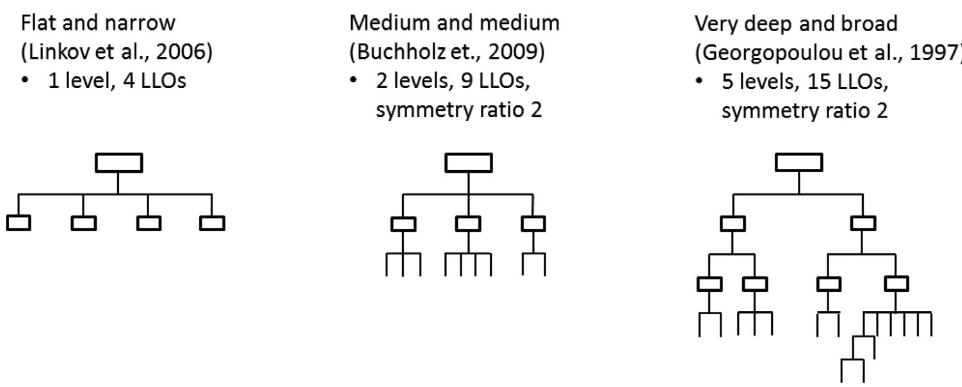


Fig. 5. Examples from the literature of three different hierarchy classes. LLO is the lowest-level objective. The references cited in this fig. are Buchholz, Rametsteiner, Volk, & Luzadis, 2009, Linkov et al., 2006, Georgopoulou, Lalas, & Papagiannakis, 1997.

The hierarchies used in the cases were divided into twelve groups in terms of their depth and breadth (see Table 4 and examples of different objectives hierarchy structures in Fig. 5). The most common depth was a hierarchy with two levels. With respect to breadth there was an approximately equal number of medium, broad and very broad hierarchies (19–20 cases in each class). The flat hierarchies were typically narrower than the deepest ones.

The symmetry of each hierarchy having more than one objective level was analysed by dividing the total number of lowest-level objectives in the largest branch by the number of lowest-level objectives in the smallest branch. A ratio of one indicates that all branches have an equal number of the lowest-level objectives, a large ratio that there is a small and a large branch in the hierarchy. Symmetry ratios varied between 1–10, the mean being 3.1 (median 2.5). Most frequently, the ratio was in the range 1–2. Only five of the 61 hierarchies were “symmetrical” having a ratio of one.

Costs was by far the most common objective; it was included in some form in 38 of the 61 cases (Fig. 6).The following cost terms were used: annual costs, capital costs, construction costs, costs of access, environmental costs, installation costs, investment costs, maintenance costs, management costs, operational costs, running costs, technology costs, transport costs and treatment costs. In thirteen cases, costs was the main objective (at level 1) without any sub-objectives, in eleven it was divided into sub-objectives, and in twelve cases it was a sub-objective below an economic objective. The second most common objective was water quality occurring in 21 cases, and the third one employment in 18 cases.

5.1.3. Methods and participants

MAVT (e.g. swing, SMART, MACBETH) was applied in 24 cases, AHP in 20 and PROMETHEE in six cases. Simos' playing card approach (Figueira & Roy, 2002) was used for weight elicitation in five cases with PROMETHEE or another outranking method (Table S2 in the SM). The results of different weighting techniques were compared in four cases (de Jalon, Iglesias, Cunningham, & Diaz, 2014; Lienert et al., 2016; Sorvari & Seppälä, 2010; Zardari, Naubi, Roslan, & Shirazi, 2014). The hierarchical weighting procedure dominated; only one article (Petersson, Giupponi, & Feas', 2007) mentions that the weights were assigned non-hierarchically. Eight articles do not mention the weighting approach and based on the information in the article we could not find out which one was applied.

The participants' preferences were gathered using interviews (26 cases), questionnaires (17), workshops (16) or combinations of these (9). For instance, Scholten et al. (2015) used a two-step procedure consisting of an online survey and a later face-to-face interview. The aim of the survey was to familiarise stakeholders and to screen irrelevant objectives before interviews. Often, the weight elicitation process was participatory and weights were defined in close cooperation with scientists, decision makers and other stake-

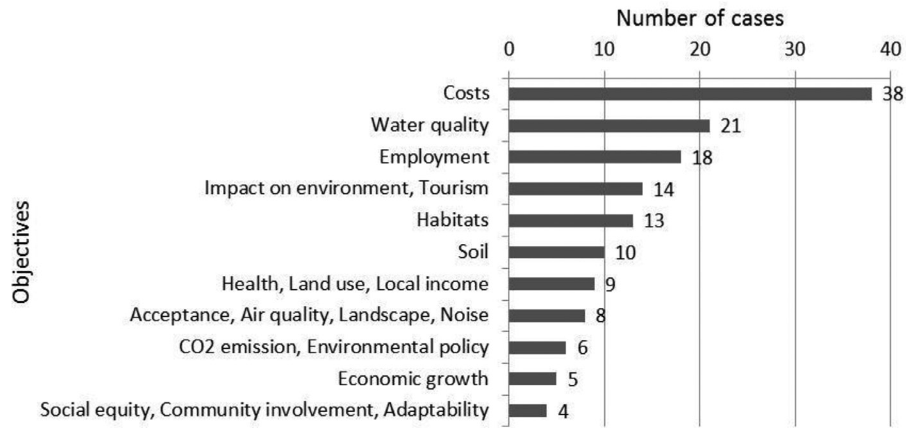


Fig. 6. Frequency of lowest-level objective types. Included are those objective types which occurred in more than three cases. Objectives in the same row have an equal number of occurrences.

Table 5
Distribution of the global weights of the lowest-level objectives (LLOs) in the four hierarchy types. Note that the weight class 0–0.1 is divided into sub-classes marked with *. n refers to the number of objective weights given in each hierarchy type.

Weight	Narrow ($n=50$) (≤ 5 LLOs)	Medium ($n=745$) (6–10 LLOs)	Broad ($n=892$) (11–15 LLOs)	Very Broad ($n=1257$) (>15 LLOs)
0–0.1	10.0	44.4	73.5	92.1
* 0–0.02	0.0	7.7	18.9	41.0
*>0.02–0.05	4.0	12.6	26.8	29.3
*>0.05–0.1	6.0	24.2	27.8	21.8
>0.1–0.2	34.0	44.3	22.6	6.0
>0.2–0.3	42.0	9.4	3.3	1.4
>0.3–0.4	4.0	1.5	0.4	0.3
>0.4–0.5	6.0	0.1	0.1	0.2
>0.5–0.6	2.0	0.3	0.0	0.1
>0.6–0.7	2.0	0.0	0.0	0.0
>0.7–0.8	0.0	0.0	0.0	0.0
Sum	100	100	100	100
Mean weight	0.26	0.12	0.07	0.04

holders (e.g. Rahman et al., 2015). In some cases participants had an opportunity to change their initial weights based on group discussions (e.g. Stefanopoulos, Yang, Gemitzi, & Tsagarakis, 2014, Straton, Jackson, Marinoni, Proctor, & Woodward, 2011).

In 33 of 61 cases weights were given by stakeholders, in seventeen they were based on experts' opinions and in three cases the source was policy makers or decision makers. It is noteworthy that the distinction between these three groups is not always unambiguous because, for instance, an expert can also be a stakeholder. In six cases, the objectives' weights were given by the authors based on the results of public surveys, for instance. The number of stakeholders who were asked to provide weights varied from one to 314 (highest in Lienert et al., 2016).

The number of weight profiles in the cases varied from one to twenty-five (highest in Karjalainen, Marttunen, Sarkki, & Rytkönen, 2013, see Fig. S1 in the supplementary material). In 30 of the 61 cases, only one weight profile was presented. The participants' weights were presented in three ways: (i) 12 cases showed individual weights, (ii) 27 cases showed mean weights of the different stakeholder groups or clusters having similar viewpoints and (iii) 20 cases showed mean weights over all participants. A clustering technique was sometimes applied to determine ideologically homogenous groupings based on the objectives' weights (e.g. Garmendia & Gamboa, 2012, Pascoe et al., 2009). Marttunen and Hämäläinen (1995) divided respondents into three groups: proponents, opponents and neutrals based on their a priori attitude towards flood protection. Mustajoki et al. (2011) identified three groups based on the rankings of the alternatives. In some cases the results of participants were finally aggregated, averaging over the individuals or giving participants a weight depending on their relative power in the decision process (Luè & Colorni, 2014).

5.1.4. Distribution of the objectives' weights

In the narrow hierarchies, 90% of the weights were higher than 0.1, whereas in the very broad hierarchies only 8% were higher than 0.1 (Table 5). The proportion of the weights higher than 0.3 was 14% and 0.6% in the narrow and very broad hierarchy, respectively. In 13 of 61 cases, at least one objective received zero weight. The total number of zero weights was 142 which is 5% of the total number of weights in the cases (2944). Three-fourths of the zero weights (101) were from three cases (Luè & Colorni, 2014; Scholten et al., 2015; Zheng et al., 2016).

5.1.5. Costs objective

Costs was the most frequently occurring objective in the selected cases. It was included in the analysis in 38 cases (corresponding to 154 weight profiles). In most weight profiles, the weight assigned to costs was less than 0.2 (Fig. 7), the mean weight was 0.14 (median 0.12) and the highest weight 0.56 (Bascetin, 2006). The weight of the costs objective was slightly higher when it was a main objective (mean 0.15, $n=127$) than a sub-objective (mean 0.10, $n=29$). In three cases, costs were not included in the weight elicitation procedure but they were compared to the results of MCDA in a two dimensional graph (Bana e costa et al., 2004; Neckles, Lyons, Guntenspergen, Shriver, & Adamowicz, 2014) or by calculating cost-to-benefit ratios of alternatives (Akash, Mamlook, & Mohsen, 1999).

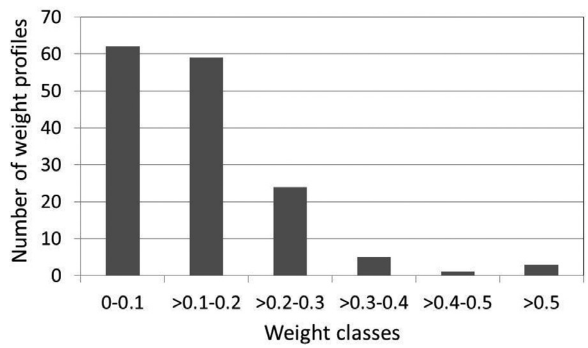


Fig. 7. Distribution of the costs weights in the weight profiles of the selected cases ($n = 154$).

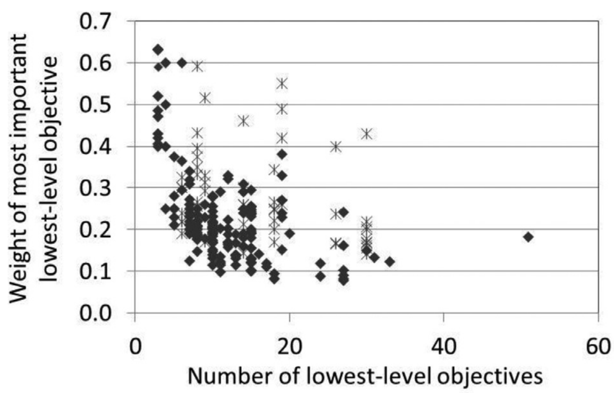


Fig. 8. Relation between the number of lowest-level objectives and the highest global weight of the lowest-level objective ($n = 230$ weight profiles). Data points which come from asymmetric hierarchies and very unevenly allocated weights are marked with *x* (see text).

5.2. Research questions

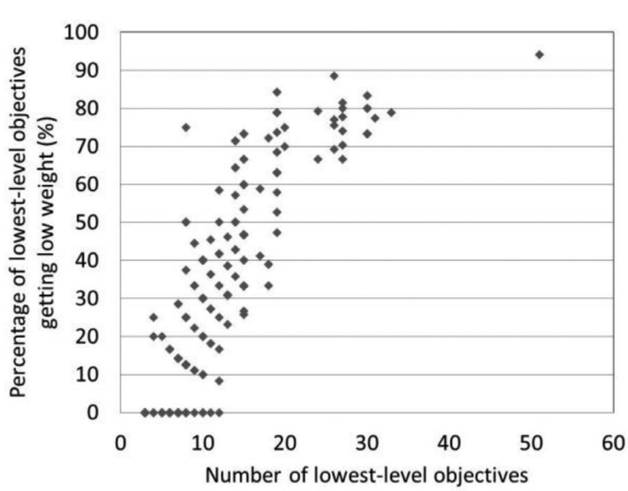
5.2.1. Research question 1: How do the size and structure of an objectives hierarchy affect weights of objectives?

Analysis 1a: There was a moderate negative correlation between the number of lowest-level objectives and the highest weight of these objectives ($r_s = -0.512$, $p < 0.001$, $n = 230$; Fig. 8). Three factors explained a large part of the variation within a case and between the cases: (i) The hierarchy branches differed in terms of their depth; e.g. one lowest-level objective was at the top level of the hierarchy and the others at the second or third hierarchy level (e.g. Kodikara, Perera, & Kularathna, 2010; Lienert, Koller, Konrad, Mcardell, & Schuwirth, 2011; Luè & Colorni, 2014). In these cases, the highest weight was typically greater than in the cases where all lowest-level objectives were at the same hierarchy level. (ii) Large difference in the number of lowest-level objectives between the hierarchy branches. For instance, in the largest analysed hierarchy (51 lowest-level objectives, Regan, Davis, Andelman, Widyanata, & Freese, 2006), the most important objective was located in the branch with 12 lowest-level objectives, whereas the largest branch had 30 lowest-level objectives. (iii) Very uneven allocation of the weights to the parental objectives. The differences in the highest weights of objectives in the four different hierarchy types (narrow, medium, broad, very broad) are presented in Fig. S2 in the SM.

Analysis 1b. The weights of objectives ≤0.05 were classified as “very low weights”. The proportion of objectives getting a weight ≤0.05 increased steeply as the number of lowest-level objectives increased from five to twenty and were strongly positively correlated ($r_s = 0.86$, $p < 0.001$, $n = 230$; Fig. 9a). In very large hierarchies having more than twenty lowest-level objectives, more than two-thirds of the objectives had weights that were ≤0.05. In smaller hierarchies, the variation in the weights was very high. The main reason for this variation was that some people distributed weights more evenly than others. This is illustrated in Fig. 9b which shows the weight distributions of three stakeholders in a transportation system case study (Luè & Colorni, 2014).

Analysis 1c: To investigate whether the number of sub-objectives within a branch has an impact on their global weights, we compared the mean weights of the sub-objectives in the smallest and largest hierarchy branches. 23 cases with asymmetric hierarchy structures were selected. From each case the hierarchy branches with the smallest and largest number of the lowest-level objectives was chosen (Fig. 10). In 17 cases the comparison was made between the lowest-level objectives located at the same hierarchy level. In six cases the lowest-level objectives located at different hierarchy levels were compared (see Fig. 1). In five of these cases the lowest-level objectives located at the top-level and second level of the objectives hierarchy, and in one case at the top-level and at the second and third level of the hierarchy.

In 71% of the weight profiles ($n = 103$) the mean global weight in the smallest hierarchy branch was higher than in the largest hierarchy branch. The difference in the mean weights was sta-



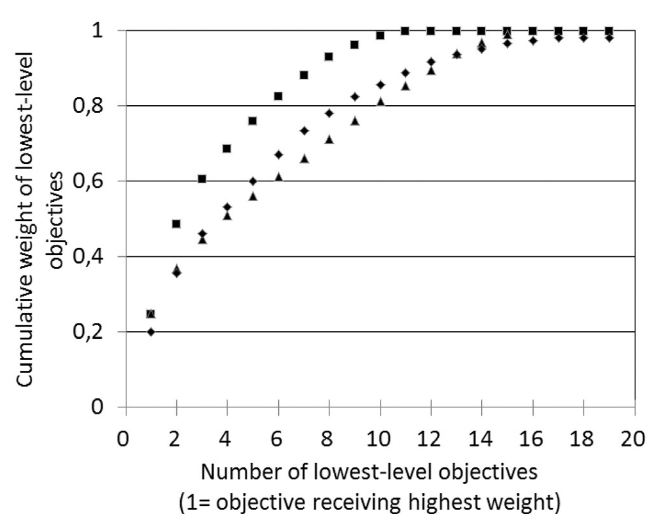


Fig. 9. (a) Percentage of lowest-level objectives having weights ≤0.05 in the analysed cases ($n = 230$ weight profiles). (b) Cumulative sum of the objectives’ weights of three stakeholders (different symbols) from one case (Luè & Colorni, 2014). Each stakeholder’s objectives are arranged in order of their weight.

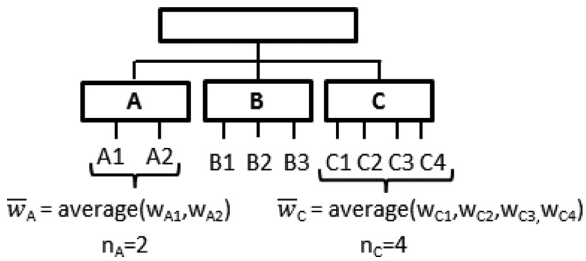


Fig. 10. Illustration of the asymmetry test (1c).

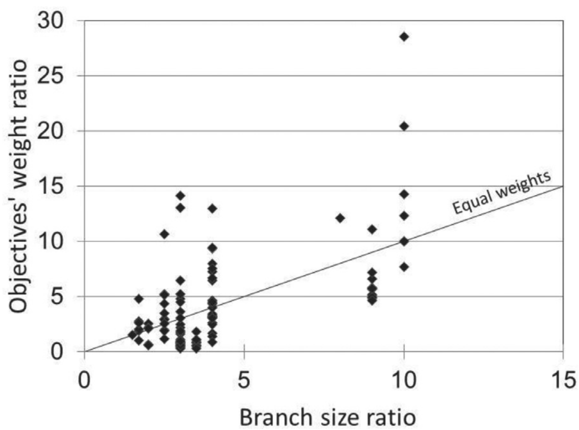


Fig. 11. Comparison of the branch size ratio and differences in mean global weights in the smallest and largest hierarchy branch (n= 103 weight profiles). x-axis: number of sub-objectives in the largest branch divided by the number of sub-objectives in the smallest branch. y-axis: sub-objectives' mean weight in the smallest hierarchy branch divided by sub-objectives' mean weight in the largest branch.

tistically significant (Related-Samples Sign Test, $Z = 23$, $p < 0.001$, $n = 103$). Branch size ratios (n_C/n_A) varied from 1.5 to 9 and the objectives' weight ratios ($\bar{w}_A/\bar{w}c$) from 0.3 to 28.5 (Figs. 10 and 11). In 32% of the weight profiles, the branch size ratio and objectives' weight ratio were rather close to each other (difference <30%) meaning that the total weight of the largest and smallest branches tends to be equal. For instance, if the number of sub-objectives in the largest branch is four times higher than in the smallest branch, then the mean weight of the sub-objectives in the smallest branch is four times higher than mean weight of sub-objectives in the largest branch.

Analysis 1d: To find out if the location of the most important objective (the lowest-level objective having the highest weight) is related to the size of the branch it is located in, we determined the size of the hierarchy branch containing the most important objective and the weights of its parental objective. Corresponding data were also determined from other branches. From this analysis we excluded cases where at least one hierarchy branch was flat (having only one hierarchy level) and compared global weights of the sub-objectives locating at the same hierarchy level. Four different situations were analysed (Table. 6). The groups 1b and 2b are of particular interest because the high ranking of the objectives belonging to the two other groups (1a and 2a) can be explained by the higher weight of the parental objective. There were 20 weight profiles where the most important objective was located in a branch whose parental objective did not have the highest weight. In 17 weight profiles (85%) this branch was the smallest one and only in three cases not. The result of Analysis 1d is in line with the result of Analysis 1c and supports the occurrence of an asymmetry bias in this sample.

Table 6
Location of the most important lowest-level objective (highest global weight, LLO) and number of weight profiles belonging to each group.

Group	Number of weight profiles
1. Most important LLO is in the smallest hierarchy branch.	
a. Weight of its parental objective is higher than weights of the other parental objectives.	25
b. Weight of its parental objective is lower than or equal to weights of the other parental objectives.	17
2. Most important LLO is not in the smallest hierarchy branch.	
a. Weight of its parental objective is higher than weights of the other parental objectives.	20
b. Weight of its parental objective is lower than or equal to weights of the other parental objectives.	3
Total	65

5.2.2. Research question 2: How are weights distributed across economic, social and environmental objectives?

Analysis 2a: Thirty-one cases and 124 weight profiles including both environmental and economic objectives were selected. The weights of environmental objectives, with a mean of 0.38 (median 0.36), were considerably higher than the weights of economic objectives (Fig. 12a), mean 0.22 (median 0.20). The difference in the mean weights was statistically significant (Related-Samples Sign Test, $Z = 28$, $p < 0.001$, $n = 124$). 77% of the participants gave higher weight to the environmental objective than to the economic objective, and for 44% the weight ratio was higher than two.

Analysis 2b. Twenty-five cases and 96 weight profiles included both social and economic objectives. The weights of social objectives, with a mean of 0.29 (median 0.28), were moderately higher than those of the economic objectives, with a mean of 0.21 (median 0.19). The difference in the mean weights was statistically significant (Related-Samples Sign Test, $Z = 28$, $p < 0.001$, $n = 96$). For 71% of the participants, the weight of the social objectives was higher, and for 35% the weight of social objectives was more than two times higher than the weight of the economic objectives (Fig. 12b).

5.2.3. Research question 3: Is there support for the equalising bias?

Analysis 3: To find out whether there is support for the equalising bias, the lowest and highest weights of top level objectives were compared. A high positive correlation would indicate that the weights are close to each other, which could be indicative of the equalising bias. However, this correlation was very weak and not statistically significant ($r_s = 0.196$, $p = 0.056$, $n = 96$). Only 5% of the weight ratios were close to equal (weight ratio >0.8–1, Fig. 13). The means of the lowest and highest weights were 0.15 and 0.43, respectively.

5.2.4. Summary of results

The main results can be summarised as follows (Table. 7):

- The higher the number of lowest-level objectives, the lower was the highest weight (Analysis 1a).
- The higher the number of lowest-level objectives, the higher was the proportion of objectives getting very low weights (Analysis 1b).
- The mean of the global weights of lowest-level objectives in the largest branches was lower than in the smallest branches (Analysis 1c). The objective having the highest global weight was located most often in the smallest branch (Analysis 1d).
- The weights of environmental and social objectives were generally higher than those of the economic objectives (Analyses 2a and b).

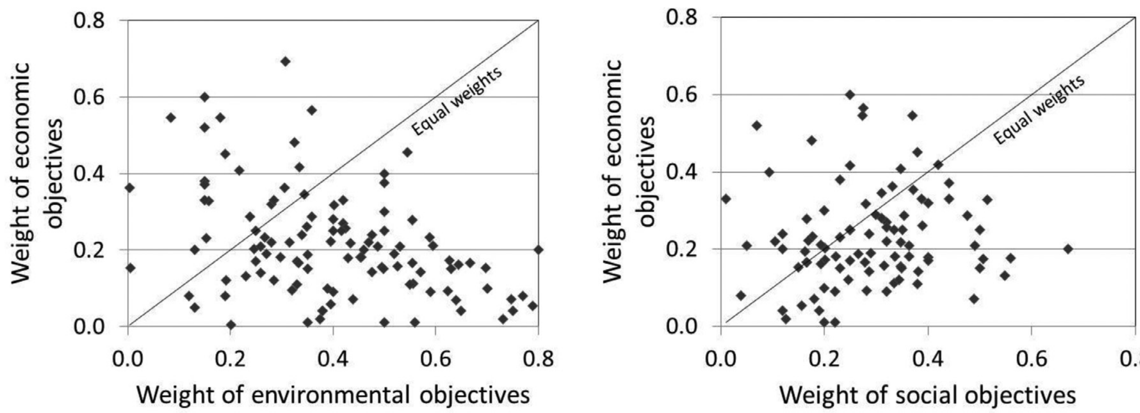


Fig. 12. (a) Comparison of the weights of environmental and economic objectives (n= 124 weight profiles). (b) Comparison of the weights of social and economic objectives (n= 96 weight profiles).

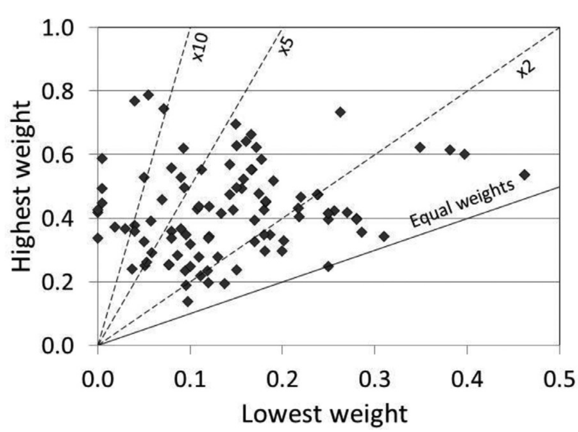


Fig. 13. Comparison of the lowest and highest weights of main objectives in each weight profile (n= 96 weight profiles). The data points are always above or on the equal weight line because in each weight profile the lowest weight is ≤ highest weight. Dashed lines illustrate where the highest weight is two, five or ten times higher than the lowest weight.

- The minimum and maximum weights of the main objectives differed substantially (Analysis 3).

These results are mostly consistent with earlier findings from the literature (presented in Section 2.2). The only exception is Analysis 3, which did not give support to the equalising bias. The results of the Analyses 1a–d are aligned with logic. The total weight is fixed and hence, as the number of objectives increases, the expected share of an individual objective decreases.

Table 7
Summary of the main results (n refers to the number of weight profiles).

	Analysis main results	Support for earlier research
1	How do the size and structure of an objectives hierarchy affect the weights of objectives?	
1a	Highest global weights of the lowest-level objectives • Narrow hierarchy (≤5 lowest-level objectives): mean 0.4 (n = 20) • Very broad hierarchy (>15 lowest-level objectives): mean 0.18 (n = 51)	Yes
1b	Proportion of lowest-level objectives getting global weight ≤0.05 • Narrow hierarchy (≤5 lowest-level objectives): 3% (n = 20) • Very broad hierarchy (>15 lowest-level objectives): 71% (n = 51)	Yes
1c	Objectives' weights in branches of different size • In 71% of the weight profiles, the objectives' mean weight was higher in the smallest branch than in the largest branch (n = 103)	Yes
1d	• In 85% of the weight profiles, the lowest-level objective receiving the highest weight was in the smallest branch (n = 65)	Yes
2	How are weights distributed across economic, social and environmental objectives?	
2a	• Environmental objectives: mean 0.38 (n = 124) • Economic objectives: mean 0.22 (n = 124)	Yes
2b	• Social objectives: mean 0.29 (n = 96) • Economic objectives: mean 0.21 (n= 96)	Yes
3	Is there support for the equalising bias? (n= 96) • Mean of main objectives' lowest weights: 0.15 • Mean of main objectives' highest weights: 0.43 • In 95% of the observations, the difference in the maximum and minimum weights of main objectives was larger than 20%	No

6. Discussion and recommendations

6.1. How does the objectives hierarchy influence the objectives' weights?

We found a strong relationship between the number of objectives and their weights (Analyses 1a and b): the higher the number of lowest-level objectives, the lower was the highest weight and the higher was the proportion of objectives which received very low weights. These results have two important consequences. First, there is a risk that the most important objective is inappropriately overridden by a large number of less important objectives. For instance, in three cases (García De Jalón, Iglesias, Cunningham, & Pérez Díaz, 2013; Lee & Chan, 2008; Rahman et al., 2015) where a very broad hierarchy (17–27 lowest-level objectives) was used, the cumulative sum of the weights of the five to eight lowest weighted objectives was equivalent to the weight of the objective getting the highest weight. It is possible that these weight allocations corresponded to the participants' real opinions, but it is just as possible that they were consequences of hierarchy related biases. Second, in all ten cases which had more than twenty lowest-level objectives, more than two-thirds of the objectives' weights were 0.05 or less. This raises the practical question whether simpler hierarchies that exclude the least important objectives would have resulted in a more efficient, understandable and meaningful decision support process. These findings have connections to the "simple heuristics" literature. Evidence suggests in certain situations simpler models perform on par or even better than more complex models in decision problems (e.g. Katsikopoulos & Fasolo, 2006, Katsikopoulos & Gigerenzer, 2013, Keller & Katsikopoulos, 2016).

This study suggests (Analyses 1c and d) that hierarchical weighting is prone to a bias which we call the asymmetry bias. It has similarities with the splitting bias but has an opposite effect on the weights. In the asymmetry bias, the higher the number of sub-objectives, the lower is each sub-objective's weight. In the splitting bias, however, the division of an objective increases its weight. The asymmetry bias occurs only in hierarchical weighting, whereas the splitting bias can occur both in hierarchical and non-hierarchical weighting. This phenomenon has received little attention in the MCDA literature. Hobbs and Meier (2000, p. 212) noted that the objectives from larger hierarchy branches tend to receive lower weights per objective than those of smaller branches.

A possible reason for the asymmetry bias is that it is cognitively very demanding to define weights taking into account both the number of sub-objectives and their ranges. For the purpose of illustration, let us consider an example with two objectives (A, B) having two (A1, A2) and five (B1–B5) sub-objectives. Let us further assume that all sub-objectives are globally of equal importance, they should thus each receive a global weight of 1/7 (0.14). To give the intended global weights to the sub-objectives, the objective B should receive a 2.5 times higher weight than the objective A.

Each of the authors has substantial experience of working as a decision analyst in a large number of real-world applications over many years (15–30). Some of our observations in practice are in accordance the findings of this study (e.g. the asymmetry bias) and have consequently influenced us to seek ways to mitigate against the problem but we had not previously been fully aware of others (e.g. the sum of the weights of a few least important objectives can equal the weight of the most important objectives).

6.2. Is the importance of economic objectives understated in MCDA?

The objectives' weights should reflect the participants' opinions of their relative importance in the decision in question. However, it is not possible to determine how well the given weights captured the participants' true preferences in the selected cases. Moreover, there is no threshold above or below which an objectives' weight could be considered as being too high or too low. Therefore, analysing only the weights of objectives cannot answer our question. Below we discuss the possible reasons why environmental objectives received much higher weights than economic objectives in the selected cases, mean weights being 0.38 and 0.22 (Analysis 2a).

Generally, there was not much discussion about the reasons for the high or low weights in the papers. Brown et al. (2001) remarked that a good environmental status forms the basis for maintaining long-term socio-economic growth, and that this connection is particularly strong in conditions where people's livelihood depends on nature. High weights for natural values due to the proximity of a national park was also mentioned (De Jalon et al., 2014).

The most important factor which affects the weights assigned to objectives is the people who are involved in the process. The preferences of stakeholders are typically strongly related to the mission of organisation they represent (see e.g. Collier, Bates, Wood, & Linkov, 2014). In several cases, a diverse group of stakeholders were engaged in the MCDA process (e.g. Karjalainen et al., 2013, Luè & Colorni, 2014, Zheng et al., 2016) and many of them were purposefully selected to cover a wide variety of perspectives. Typically, many were especially concerned about the environmental and social impacts of the projects. Additionally, only few stakeholders were those who would have to carry the costs if the project was realised. In our study, the mean weights of environmental, social and economic objectives were calculated across all participants, and it is likely that "overpresentation" of participants stressing environmental and social objectives may have increased the weights of these objectives.

The number of economic objectives (mean 3.1) was lower than the number of environmental (5.7) and social objectives (3.8). This smaller number might also partly explain the smaller weight of economic objectives. However, in hierarchical weighting, which was used in most cases, the weights of the main objectives define how the weights are allocated to sub-objectives. Thereby, in hierarchical weighting the number of sub-objectives does not relate as directly to the weights as in non-hierarchical weighting, where the lowest-level objectives are weighted directly and define the higher level weights.

Weight elicitation questions which do not explicitly acknowledge the impact ranges implied by the scales used to describe the alternatives' performances can lead to weights which do not reflect the participants' true opinions. However, it was not possible to study this topic for the reasons mentioned above. Two examples illustrate the problems which can be encountered if impact ranges are not considered: Silva, Morais, and Almeida (2010), who used a direct rating technique, report that a low weight on economic objectives resulted in surprises in the alternatives' rankings; the most expensive alternative was ranked first, which was not in line with the expectations of two stakeholders. In a case concerning wind farms, in which Simos' card approach was applied in the weight elicitation (Polatidis & Morales, 2014), one stakeholder gave a seven times higher weight to one unit change in public acceptance (range 4.5–5.5 on a scale of 1–10) compared to 117 million Euros difference in investment costs.

The lower weights assigned to economic objectives can also be a consequence of small difference in alternatives' impact ranges. Lienert et al. (2016) noted that higher costs of a good wastewater disposal system were not substantial enough to justify trade-offs with the negative environmental effects of a cheaper but worse wastewater system. In other cases, this might be caused by a real bias.

Taboo-tradeoffs (e.g. Tetlock, Kristel, Elson, Green, & Lerner, 2000), also called protected values (e.g. Baron & Spranca, 1997), may also explain why environmental objectives receive higher weights than economic objectives. People having such values are not willing to make trade-offs with other values, particularly economic values, and are probably more likely to be insensitive to consequences (Baron & Spranca, 1997). The result of giving a higher priority to environmental and social objectives is also in line with observations of Keeney (2002) and Gregory et al. (2012) who state that environmental and social objectives receive higher weights because they are generally considered morally and ethically more important than economic objectives.

The participants´unfamiliarity with large amounts of money can also partly explain lower weights assigned to economic objectives (Weber & Borcherding, 1993). Moreover, the used attributes may influence the weights. If environmental or social impacts are described with proxy attributes, people may have to make judgments of their consequences without support or a full understanding and this can lead to overweighting as demonstrated by Fischer et al. (1987).

An open question is whether the hierarchical weighting procedure can result in overweighting of environmental and social objectives. Impact ranges easily become ill-defined and comparisons complex at the upper levels of the hierarchy (e.g. Stillwell et al., 1987). Consequently, the labelling of the objectives might have a greater influence on the weights at the top level than at the lowest-level of the hierarchy.

6.3. Do people give equal weights?

The equalising bias is suggested to be a result of the anchor-and-adjust heuristic, meaning that the participant starts with an equal allocation of weights among attributes and then adjusts the

weights to reflect his or her innate preferences (Kahneman, Slovic, & Tversky, 1982). As a result, the participant's weights within a group of compared objectives can be more similar to one another than if this cognitive strategy is not applied (Jacobi & Hobbs, 2007). We explored the highest and lowest weights of the main objectives to determine whether their ratio would support the equalising bias (Test 3). However, we did not find any indication of the occurrence of this bias (see Table. 7).

We also analysed whether the tendency to give equal weights is higher if there is a so called three pillar structure, i.e. if the common sustainability objectives, economic, social and environmental, are located at the top level of the hierarchy. It could be intuitively attractive and easy to give equal weights to all of these objectives. For MCDA, it has even been suggested that these three dimensions could be considered equally important (Munda, 2006). However, this does not take into account the actual impact ranges. We analysed 12 such cases and found that only in five of the 55 weight profiles was the ratio of the lowest weight to the highest weight higher than 0.75.

It is possible that the tendency to give equal weights is stronger if people have a weak interest in the decision in question and no clear preferences, which might especially apply to student experiments (Pöyhönen & Hämäläinen, 2000). It is also possible that some weighting methods (e.g. point allocation) are more prone to the equalising biases than others. Evidence for the equalising bias comes mainly from resource allocation problems and probability assessments (e.g. Fasolo, Morton, & Von Winterfeldt, 2011) where a participant has to assign available resources to various uses or allocate the total probability of one to different events. Maybe this bias is less applicable to MCDA-type of weight distributions, but this should be further researched in future.

In MCDA the equalising bias can be a real problem because the weights should, according to theory, reflect the differences in the alternatives' impacts as well as in the perceived "importance". Particularly in cases where there are large differences in the impact ranges over the objectives, the equalising bias can greatly distort the results. Giving equal weights may be an indication that person has not understood the meaning of weight or is not able or willing to evaluate the objectives' differences.

6.4. Recommendations

Constructing the objectives hierarchy is often an activity which requires significant commitment of stakeholders and experts. It would be unfortunate if this effort is undermined by hierarchy structure induced biases in the weight elicitation process. Below, we present general recommendations based on the findings of this study. However, each case has its own purposes and characteristics and, therefore, these recommendations should not be followed too literally.

- Build concise objectives hierarchies (relates to analyses 1a and b). Although it is difficult to give a precise recommendation concerning the size of hierarchy, this study and our earlier experiences suggest that if the number of objectives in the initial objective hierarchy exceeds 15, then opportunities to simplify the hierarchy should be carefully considered.
- Carefully consider if asymmetric hierarchies are appropriate and in that case use either weighting procedures which are insensitive to the hierarchy structure or consistency check questions across branches (relates to analyses 1c and d). This can help to avoid the asymmetry bias, which seems especially relevant for the hierarchical weighting procedure. Hämäläinen and Alaja (2008) suggest constructing symmetric/balanced hierarchies to avoid the splitting bias and this is illustrated by Lienert et al. (2016).
- Avoid deep hierarchies because they are more prone to behavioural and procedural biases than flatter hierarchies (relates to analyses 1c,d, 2a,b). The more hierarchy levels there are, the more trade-offs have to be made at the different levels in hierarchical weighting. At the higher level of the hierarchy trade-offs are cognitively more demanding and prone to the range insensitivity bias (von Nitzsch & Weber, 1993). In deeper hierarchies, there are also more hierarchy levels which may have a different number of sub-objectives which increases the risk of the asymmetry bias.
- Consider different ways to include costs in the multicriteria evaluation to find the most appropriate one for the decision in question (relates to analyses 2a and b). Possible options are, for instance: (i) include costs as a main objective or as a sub-objective. As location in the hierarchy might greatly affect the allocated weight, this decision is worth careful consideration; (ii) develop a hierarchy which has a "costs" branch (including monetary and non-monetary impacts) and a benefits branch and use sensitivity analysis to analyse the impacts of different weights of these branches. Alternatively, with this structure it is not necessary to assign weights to the costs and benefits objectives as the trade-offs can be visualised in a cost-benefit graph to highlight "efficient" options and allow decision makers to consider where they prefer to be on the efficient frontier; or, similarly, (iii) compare the alternatives' costs and MCDA overall values (non-monetary costs included) in a two dimensional figure.
- Compare the highest and lowest weights of the objectives across hierarchy in a hierarchical weighting procedure (relates to analyses 1a and b). The comparison is particularly important when the number of objectives is high and when there are several hierarchy branches which have a different number of sub-objectives.
- Use different techniques to diminish the risk of mistakes and biases. Highly recommendable techniques to improve the quality of weight elicitation are: (i) using interactive and iterative weight elicitation procedures which enable asking for arguments for the weights and can help to detect inconsistencies (Marttunen & Hämäläinen, 2008; Schuwirth et al., 2012; von Winterfeldt & Fasolo, 2009); (ii) asking consistency check questions during weight elicitation (e.g. Lienert et al., 2011, Montibeller & von Winterfeldt, 2015); and (iii) training and educating participants about the weight elicitation procedure and different types of biases (e.g. Anderson & Clemen, 2013, Hämäläinen & Alaja, 2008).
- The weighting technique, whether hierarchical or non-hierarchical, determines whether the variations of the lowest-level or top level objectives have a greater impact on the weights[3]. In non-hierarchical weighting, merging or dividing the lowest-level objectives can influence their weights (the splitting bias, e.g. Borcherding & von Winterfeldt, 1988, Weber et al., 1988). In hierarchical weighting, variations in the weights at the top level are more important. A hierarchical weighting procedure (top-down) was used in most cases. However, in this method determining trade-offs can be very demanding at the intermediate and top levels of the objectives hierarchy and the method is prone to the range insensitivity and the asymmetry

[3] The lowest-level objectives' weights are central because they are used to calculate the overall values of alternatives. In hierarchical weighting, the lowest-level objective weights are calculated by multiplying their local weights with the local weights of their parental objectives. In non-hierarchical weighting intermediate and top level weights are not necessary because the lowest-level weights are assigned directly by comparing the lowest-level objectives to each other. However, presentation of these higher-level weights can illustrate how weights are distributed between the main objectives and can also support a sensitivity analysis.

bias. In “theory” a non-hierarchical approach would appear to be less subject to these biases. However, there is an insufficient number of published cases to allow an exploratory comparison of approaches.

We believe that the results of this study are most useful in complex cases which may lead to the development of large hierarchies. Developing more concise and well-structured hierarchies can improve communication among participants, leading to less laborious weight elicitation and less costly processes. Taking into account the recommendations presented here can also lead to objectives' weights which better reflect participants' true opinions.

6.5. Reflections on the use of meta-analysis in this study, future potential and limitations

MCDA applications in environmental decision making are typically time-consuming; a large case involving 10–15 stakeholders can easily take 6–18 months. Thus, it is difficult for a research or consultancy group to gather enough data to allow for extensive statistical analyses based only on their own cases. We found meta-analysis to be a cost-efficient research method; it took only 5 months to identify the cases, analyse the data and write this article. Meta-analysis was powerful tool to broaden our perspective from the individual case to a “bird's-eye view” of a range of applications. The analysis gave a good overall understanding of the general patterns in weight distributions. It also generated new insights showing that asymmetry in hierarchy branches can greatly affect the weights of objectives.

Although the results of this study are mostly in parallel with earlier laboratory experiments and with our observations in real-world applications, caution should be exercised in generalising from the results, for the following reasons: the sample size is relatively limited (61 cases), it is a purposive (judgement) sample selected to ensure representation of a range of hierarchy sizes, and there is potential bias due to the overpresentation of Swiss and Finnish cases. It is possible that overpresentation of our cases may have brought some systematic facilitator induced bias in the weight elicitation. However, the authors of this paper worked as decision analysts only in two cases, which of course diminishes the risk that a specific systematic facilitator dependent error occurs in all these cases.

The number of weight profiles in the selected cases varied from one to twenty-five meaning that the latter contribute 25 times more data points to the statistical analyses. If there were case related systematic biases, for instance due to the behaviour of a single decision analyst, these cases would influence the results of the meta-analysis most strongly.

A challenge faced was to understand and interpret the wide variations in the results of the meta-analysis. Some of the variation could be explained by differences in the hierarchy structure and how evenly the participants' weights were distributed across the objectives. Part of the variation may be due to the different weight elicitation techniques or the different levels of interaction between the analyst and participants in the weight elicitation process. Another possible reason is that some cases present individual weights, whereas others show mean weights of a larger group of people. It was not possible to identify the impacts of these factors on the weights.

Furthermore, it was not possible to study all behavioural and procedural biases related to the structure and size of objectives hierarchies. For instance, the splitting bias (i.e. division of one objective into two or more objectives) can more conveniently be investigated in experiments, in which either the same participants assign weights to each of two different objectives hierarchies (e.g. Hämäläinen & Alaja, 2008) or a large number of people use different hierarchies in which some objectives are split differently (e.g. Borcherding & von Winterfeldt, 1988). The same applies to the range insensitivity bias (Fischer, 1995).

This study indicates that, in addition to controlled experimental research related to cognitive and motivational biases suggested by Montibeller and von Winterfeldt (2015), other types of research, including meta-analysis, can also contribute to the research on biases. We see this research as a first step in the use of meta-analysis in MCDA. It would be interesting to extend such a meta-analysis from environmental and energy applications to other domains. In addition, comparison of weight distributions derived from different preference collection methods (e.g. face-to-face interviews, surveys) and different weighting methods (e.g. swing, trade-off, AHP) could produce information which might be useful to improve current practices.

We are more aware, in retrospect, of limitations of our meta-analysis. To some extent these could be addressed in two ways. From the researchers' part the potential for more systematic selection of cases should be considered. However, the number and nature of publications relating to MCDA in practice, in particular the diversity of published detail, is such that a large scale random selection of cases for analysis is unlikely to be possible and purposive sampling will always be necessary. The potential for such analyses could also be facilitated by the authors of cases studies for publication, perhaps with the encouragement of the relevant academic/practitioner community and associated journal editorial boards. It was our experience that although there is a large number of published MCDA applications, the number of cases which were appropriate for a meta-analysis was much smaller due to limited documentation of the weight elicitation procedures and of the weights assigned to objectives by participants. More systematic documentation of processes and collected data (e.g. in supplementary materials) would improve the opportunities to carry out large-scale meta-analyses.

7. Conclusions

The aim of this study was twofold. First, we used a sample of MCDA applications to investigate whether prior findings regarding objectives hierarchy related biases, mostly based on laboratory experiments, can be found in real-world applications. Second, we examined the benefits and pitfalls of meta-analysis in the context of this study. The main results are: (i) the number of objectives influences both the highest and lowest weights assigned to objectives, (ii) the asymmetry of an objectives hierarchy influences the weights of objectives when using a hierarchical weighting procedure, (iii) environmental objectives received considerably higher weights than economic objectives in the selected cases and (iv) no evidence for the equalising bias was found. This study is the first of this type of meta-analysis in MCDA and we hope that its publication will encourage further studies which seek to promote learning from practical applications of MCDA and to extend the methodology in this context through the use of larger samples and potentially more sophisticated approaches to analysis.

Supplementary material

Supplementary material associated with this article can be found, in the online version at doi:10.1016/j.ejor.2017.02.038.

Acknowledgements

We thank Melanie Graf for her help in collecting and analysing the data and Pascal Bücheler for the technical assistance. Alice

Aubert, Fridolin Haag, Jyri Mustajoki and Raimo P. Hämäläinen gave insightful comments to the manuscript. We thank three reviewers for their insightful comments and suggestions. We are grateful for the Directorate of the Swiss Federal Institute of Aquatic Science and Technology (Eawag) for research funding.

References

Akash, B. A., Mamlook, R., & Mohsen, M. S. (1999). Multi-criteria selection of electric power plants using analytical hierarchy process. *Electric Power Systems Research, 52*, 29–35. doi:10.1016/s0378-7796(99)00004-8.

Ananda, J., & Herath, G. (2009). A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics, 68*, 2535–2548.

Anderson, R. M., & Clemen, R. (2013). Toward an improved methodology to construct and reconcile decision analytic preference judgments. *Decision Analysis, 10*, 121–134. doi:10.1287/deca.2013.0268.

Antunes, P., Karadzic, V., Santos, R., Beca, P., & Osann, A. (2011). Participatory multi-criteria analysis of irrigation management alternatives: The case of the Caia irrigation district, Portugal. *International Journal of Agricultural Sustainability, 9*, 334–349. doi:10.1080/14735903.2011.582358.

Arnqvist, G., & Wooster, D. (1995). Meta-analysis: Synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution, 10*, 236–240.

Balasubramaniam, A., & Voulvoulis, N. (2005). The appropriateness of multicriteria analysis in environmental decision-making problems. *Environmental Technology, 26*, 951–962. doi:10.1080/09593332608618484.

Bana e costa, C. A., Da silva, P. A., & Correia, F. N. (2004). Multicriteria evaluation of flood control measures: The case of Ribeira do Livramento. *Water Resources Management, 18*, 263–283. doi:10.1023/B:WARM.0000043163.19531.6a.

Bardolet, D., Fox, C. R., & Lovallo, D. (2011). Corporate capital allocation: A behavioral perspective. *Strategic Management Journal, 32*, 1465–1483. doi:10.1002/smj.966.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16. doi:10.1006/obhd.1997.2690.

Bascetin, A. (2006). A decision support system using analytical hierarchy process (AHP) for the optimal environmental reclamation of an open-pit mine. *Environmental Geology, 52*, 663–672. doi:10.1007/s00254-006-0495-7.

Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function. *European Journal of Operational Research, 26*, 7–21.

Belton, V., & Stewart, T. J. (2002). *Multiple criteria decision analysis - An integrated approach*. Boston: Kluwer.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American Economic Review, 91*, 79–98. doi:10.1257/aer.91.1.79.

Borcherding, K., & Von winterfeldt, D. (1988). The effect of varying value trees on multiattribute evaluations. *Acta Psychologica, 68*, 153–170.

Bottomley, P. A., & Doyle, J. R. (2001). A comparison of three weight elicitation methods: Good, better, and best. *Omega-International Journal of Management Science, 29*, 553–560. doi:10.1016/S0305-0483(01)00044-5.

Brans, J. P., Vincke, P., & Mareschal, B. (1986). How to select and how to rank projects - The PROMETHEE method. *European Journal of Operational Research, 24*, 228–238. doi:10.1016/0377-2217(86)90044-5.

Brown, K., Adger, W. N., Tompkins, E., Bacon, P., Shim, D., & Young, K. (2001). Trade-off analysis for marine protected area management. *Ecological Economics, 37*, 417–434. doi:10.1016/S0921-8009(00)00293-7.

Buchholz, T., Rametsteiner, E., Volk, T. A., & Luzadis, V. A. (2009). Multi Criteria Analysis for bioenergy systems assessments. *Energy Policy, 37*, 484–495. doi:10.1016/j.enpol.2008.09.054.

Collier, Z. A., Bates, M. E., Wood, M. D., & Linkov, I. (2014). Stakeholder engagement in dredged material management decisions. *Science of the Total Environment, 496*, 248–256. doi:10.1016/j.scitotenv.2014.07.044.

Cook, R. L., & Stewart, T. R. (1975). Comparison of 7 methods for obtaining subjective descriptions of judgmental policy. *Organizational Behavior and Human Performance, 13*, 31–45. doi:10.1016/0030-5073(75)90004-5.

De jalon, S. G., Iglesias, A., Cunningham, R., & Diaz, J. I. P. (2014). Building resilience to water scarcity in southern Spain: A case study of rice farming in Doana protected wetlands. *Regional Environmental Change, 14*, 1229–1242. doi:10.1007/s10113-013-0569-5.

Doyle, J. R., Green, R. H., & Bottomley, P. A. (1997). Judging relative importance: Direct rating and point allocation are not equivalent. *Organizational Behavior and Human Decision Processes, 70*, 65–72. doi:10.1006/obhd.1997.2694.

Edwards, W. (1977). How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions On Systems, Man, And Cybernetics, 7*, 326–340.

Eisenführ, F., Weber, M., & Langer, T. (2010). *Rational decision making*. Springer: Berlin.

Fasolo, B., Mcclelland, G. H., & Todd, P. M. (2007). Escaping the tyranny of choice: When fewer attributes make choice easier. *Marketing Theory Articles, 7*, 13–26. doi:10.1177/1470593107073842.

Fasolo, B., Morton, A., & Von winterfeldt, D. (2011). Behavioural issues in portfolio decision analysis. In A. SALO, J. KEISLER, & A. MORTON (Eds.), *Portfolio decision analysis: Improved methods for resource allocation*. Springer-Verlag.

Figueira, J., & Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research, 139*, 317–326. doi:10.1016/S0377-2217(01)00370-8.

Fischer, G. (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes, 62* pp. 352–266.

Fischer, G. W., Damodaran, N., Laskey, K. B., & Lincoln, D. (1987). Preferences for proxy attributes. *Management Science, 33*, 198–214. doi:10.1287/mnsc.33.2.198.

Fox, C. R., & Clemen, R. T. (2005). Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Science, 51*, 1417–1432. doi:10.1287/mnsc.1050.0409.

Franco, L. A., & Hämäläinen, R. P. (2016). Behavioural operational research: Returning to the roots of the OR profession. *European Journal of Operational Research, 249*, 791–795. doi:10.1016/j.ejor.2015.10.034.

García de jalón, S., Iglesias, A., Cunningham, R., & Pérez díaz, J. I. (2013). Building resilience to water scarcity in southern Spain: A case study of rice farming in Doñana protected wetlands. *Regional Environmental Change, 14*, 1229–1242. doi:10.1007/s10113-013-0569-5.

Garmendia, E., & Gamboa, G. (2012). Weighting social preferences in participatory multi-criteria evaluations: A case study on sustainable natural resource management. *Ecological Economics, 84*, 110–120. doi:10.1016/j.ecolecon.2012.09.004.

Georgopoulou, E., Lalas, D., & Papagiannakis, L. (1997). A multi-criteria decision aid approach for energy planning problems: The case of renewable energy option. *European Journal of Operational Research, 103*, 38–54.

Gregory, R., Failing, L., Harstone, M., Long, G., Mcdaniels, T., & Ohlson, D. (2012). *Structured decision making: A practical guide to environmental management choices*. Chichester: Wiley-Blackwell.

Hajkowicz, S., & Collins, K. (2007). A review of multiple criteria analysis for water resource planning and management. *Water Resources Management, 21*, 1553–1566. doi:10.1007/s11269-006-9112-5.

Hämäläinen, R. P. (2015). Behavioural issues in environmental modelling - The missing perspective. *Environmental Modelling & Software, 73*, 244–253. doi:10.1016/j.envsoft.2015.08.019.

Hämäläinen, R. P., & Alaja, S. (2008). The threat of weighting biases in environmental decision analysis. *Ecological Economics, 68*, 556–569. doi:10.1016/j.ecolecon.2008.05.025.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hobbs, B. F., & Meier, P. (2000). *Energy decisions and the environment: A guide to the use of multicriteria methods*. Boston: Kluwer Academic Publishers.

Huang, I. B., Keisler, J., & Linkov, I. (2011a). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *The Science of the Total Environment, 409*, 3578–3594. doi:10.1016/j.scitotenv.2011.06.022.

Huberman, G., & Jiang, W. (2006). Offering versus choice in 401(k) plans: Equity exposure and number of funds. *Journal of Finance, 61*, 763–801. doi:10.1111/j.1540-6261.2006.00854.x.

Jacobi, S. K., & Hobbs, B. F. (2007). Quantifying and mitigating the splitting bias and other value tree-induced weighting biases. *Decision Analysis, 4*, 194–210. doi:10.1287/deca.1070.0100.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge; New York: Cambridge University Press.

Karjalainen, T. P., Marttunen, M., Sarkki, S., & Rytkönen, A. M. (2013). Integrating ecosystem services into environmental impact assessment: An analytic-deliberative approach. *Environmental Impact Assessment Review, 40*, 54–64. doi:10.1016/j.eiar.2012.12.001.

Katsikopoulos, K. V., & Fasolo, B. (2006). New tools for decision analysts. *IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 36*, 960–967. doi:10.1109/Tsmca.2006.871798.

Katsikopoulos, K. V., & Gigerenzer, G. (2013). Behavioral operations management: a blind spot and a research program. *Journal of Supply Chain Management, 49*, 3–7. doi:10.1111/j.1745-493x.2012.03285.x.

Keeney, R. (2007). Developing objectives and attributes. In W. EDWARDS, R. F. MILES, & D. VON WINTERFELDT (Eds.), *Advances in Decision Analysis*. New York: Cambridge University Press.

Keeney, R. L. (2002). Common mistakes in making value trade-offs. *Operations Research, 50*, 935–945. doi:10.1287/opre.50.6.935.357.

Keeney, R. L., & Gregory, R. S. (2005). Selecting attributes to measure the achievement of objectives. *Operations Research, 53*, 1–11. doi:10.1287/opre.1040.0158.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives preferences and value tradeoffs*. New York a.o.: Wiley.

Keisler, J., & Linkov, I. (2014). Environment models and decisions. *Environment Systems and Decisions, 34*, 369–372. doi:10.1007/s10669-014-9515-4.

Keller, N., & Katsikopoulos, K. V. (2016). On the role of psychological heuristics in operational research; and a demonstration in military stability operations. *European Journal of Operational Research, 249*, 1063–1073. doi:10.1016/j.ejor.2015.07.023.

Kodikara, P. N., Perera, B. J. C., & Kularathna, M. D. U. P. (2010). Stakeholder preference elicitation and modelling in multi-criteria decision analysis – A case study on urban water supply. *European Journal of Operational Research, 206*, 209–220. doi:10.1016/j.ejor.2010.02.016.

Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton: Princeton University Press.

Lee, G. K. L., & Chan, E. H. W. (2008). The Analytic Hierarchy Process (AHP) approach for assessment of urban renewal proposals. *Social Indicators Research, 89*, 155–168. doi:10.1007/s11205-007-9228-x.

Lienert, J., Duygan, M., & Zheng, J. (2016). Preference stability over time with multiple elicitation methods to support wastewater infrastructure decision-making. *European Journal of Operational Research, 253*, 746–760. doi: hTtp://dx.doi.org/10.1016/j.ejor.2016.03.010 .

Lienert, J., Koller, M., Konrad, J., Mcardell, C. S., & Schuwirth, N. (2011). Multiple-criteria decision analysis reveals high stakeholder preference to remove pharmaceuticals from hospital wastewater. *Environment Science Technology, 45*, 3848–3857. doi:10.1021/es1031294.

Lindhjem, H., & Navrud, S. (2008). How reliable are meta-analyses for international benefit transfers. *Ecological Economics, 66*, 425–435. doi:10.1016/j.ecolecon.2007.10.005.

Linkov, I., Satterstrom, F. K., Kiker, G., Seager, T. P., Bridges, T., Gardner, K. H., & et al. (2006). Multicriteria decision analysis: A comprehensive decision approach for management of contaminated sediments. *Risk Analysis, 26*, 61–78. doi:10.1111/j.1539-6924.2006.00713.x.

Luè, A., & Colorni, A. (2014). Conflict analysis for environmental impact assessment: A case study of a transportation system in a tourist area. *Group Decision and Negotiation, 24*, 613–632. doi:10.1007/s10726-014-9403-9.

Macharis, C., Springael, J., De brucker, K., & Verbeke, A. (2004). PROMETHEE and AHP: The design of operational synergies in multicriteria analysis. Strengthening PROMETHEE with ideas of AHP. *European Journal of Operational Research, 153*, 307–317. doi:10.1016/S0377-2217(03)00153-X.

Marttunen, M., & Hämäläinen, R. P. (1995). Decision analysis interviews in environmental impact assessment. *European Journal of Operational Research, 87*, 551–563. doi:10.1016/0377-2217(95)00229-4.

Marttunen, M., & Hämäläinen, R. P. (2008). The Decision Analysis Interview Approach in the collaborative management of a large regulated water course. *Environmental Management, 42*, 1026–1042. doi:10.1007/s00267-008-9200-9.

Marttunen, M., Mustajoki, J., Dufva, M., & Karjalainen, T. (2015). How to design and realize participation of stakeholders in MCDA processes? A framework for selecting an appropriate approach. *EURO Journal on Decision Processes, 3*, 187–214. doi:10.1007/s40070-013-0016-3.

Matarazzo, & Nijkamp (1997). Meta-analysis for comparative environmental case studies: methodological issues. *International Journal of Social Economics, 24*(7/8/9), 799–811. doi:10.1108/03068299710178865.

Meline, T. (2006). Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders, 33*, 21–27.

Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management, 230*, 1–22. doi:10.1016/j.foreco.2006.03.023.

Montibeller, G., & Von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk Analysis. *Risk Analysis, 35*, 1230–1251. doi:10.1111/risa.12360.

Munda, G. (2006). Social multi-criteria evaluation for urban sustainability policies. *Land Use Policy, 23*, 86–94. doi:10.1016/j.landusepol.2004.08.012.

Mustajoki, J., Saarikoski, H., Marttunen, M., Ahtikoski, A., Hallikainen, V., Helle, T., et al. (2011). Use of decision analysis interviews to support the sustainable use of the forests in Finnish Upper Lapland. *Journal of Environmental Management, 92*, 1550–1563. doi:10.1016/j.jenvman.2011.01.007.

Neckles, H. A., Lyons, J. E., Guntenspergen, G. R., Shriver, W. G., & Adamowicz, S. C. (2014). Use of structured decision making to identify monitoring variables and management priorities for salt marsh ecosystems. *Estuaries and Coasts, 38*, 1215–1232. doi:10.1007/s12237-014-9822-5.

Nelson, J., & Kennedy, P. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental & Resource Economics, 42*, 345–377. doi:10.1007/s10640-008-9253-5.

Pascoe, S., Proctor, W., Wilcox, C., Innes, J., Rochester, W., & Dowling, N. (2009). Stakeholder objective preferences in Australian Commonwealth managed fisheries. *Marine Policy, 33*, 750–758. doi:10.1016/j.marpol.2009.02.008.

Petersson, E., Giupponi, C., & Feas', J. (2007). Decision support for strategic water management: Mdss in the large dam context. *Water International, 32*, 265–279.

Petitti, D. (1994). *Meta-analysis, decision analysis, and cost-effectiveness analysis*. New York: Oxford University Press.

Polatidis, H., & Morales, J. B. (2014). Increasing the applicability of wind power projects via a multi-criteria approach: Methodology and case study. *International Journal of Sustainable Energy*, 1–16. doi:10.1080/14786451.2014.975130.

Pöyhönen, M. (1998). *On attribute weighting in value trees*. Helsinki: University of Technology Doctoral thesis.

Pöyhönen, M., & Hämäläinen, R. P. (1998). Notes on the weighting biases in value trees. *Journal of Behavioral Decision Making, 11*, 139–150. doi:10.1002/(SICI)1099-0771(199806).

Pöyhönen, M., & Hämäläinen, R. P. (2000). There is hope in attribute weighting. *Infor, 38*, 272–282.

Pöyhönen, M., & Hämäläinen, R. P. (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research, 129*, 569–585. doi:10.1016/S0377-2217(99)00467-1.

Pöyhönen, M., Vrolijk, H., & Hämäläinen, R. P. (2001). Behavioral and procedural consequences of structural variation in value trees. *European Journal of Operational Research, 134*, 216–227. doi:10.1016/S0377-2217(00)00255-1.

Rahman, M. A., Jaumann, L., Lerche, N., Renatus, F., Buchs, A. K., Gade, R., et al. (2015). Selection of the best inland waterway structure: A multicriteria decision analysis approach. *Water Resources Management, 29*, 2733–2749. doi:10.1007/s11269-015-0967-1.

Regan, H. M., Davis, F. W., Andelman, S. J., Widyanata, A., & Freese, M. (2006). Comprehensive criteria for biodiversity evaluation in conservation planning. *Biodiversity and Conservation, 16*, 2715–2728. doi:10.1007/s10531-006-9100-3.

Reichert, P., Langhans, S. D., Lienert, J., & Schuwirth, N. (2015). The conceptual foundation of environmental decision support. *Journal of Environmental Management, 154*, 316–332. doi:10.1016/j.jenvman.2015.01.053.

Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision, 31*, 49–73. doi:10.1007/Bf00134132.

Saad, G., & Russo, J. E. (1996). Stopping criteria in sequential choice. *Organizational Behavior and Human Decision Processes, 67*, 258–270. doi:10.1006/obhd.1996.0078.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.

Salo, A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the Analytic Hierarchy Process. *Journal of Multi-Criteria Decision Analysis, 6*, 309–319.

Sayeki, Y., & Vesper, K. H. (1971). Allocation of importance and goal-dependent utility. *Scientific Interim*. Seattle Washington: University of Washington. Department of Psychology.

Schoemaker, P. J. H., & Waid, C. C. (1982). An experimental comparison of different approaches to determining weights in additive utility-models. *Management Science, 28*, 182–196. doi:10.1287/mnsc.28.2.182.

Scholten, L., Schuwirth, N., Reichert, P., & Lienert, J. (2015). Tackling uncertainty in multi-criteria decision analysis - An application to water supply infrastructure planning. *European Journal of Operational Research, 242*, 243–260. doi:10.1016/j.ejor.2014.09.044.

Schuwirth, N., Reichert, P., & Lienert, J. (2012). Methodological aspects of multi-criteria decision analysis for policy support: A case study on pharmaceutical removal from hospital wastewater. *European Journal of Operational Research, 220*, 472–483. doi:10.1016/j.ejor.2012.01.055.

Silva, V. B. S., Morais, D. C., & Almeida, A. T. (2010). A multicriteria group decision model to support watershed committees in Brazil. *Water Resources Management, 24*, 4075–4091. doi:10.1007/s11269-010-9648-2.

Sorvari, J., & Seppälä, J. (2010). A decision support tool to prioritize risk management options for contaminated sites. *Science of the Total Environment, 408*, 1786–1799. doi:10.1016/j.scitotenv.2009.12.026.

Stefanopoulos, K., Yang, H., Gemitzi, A., & Tsagarakis, K. P. (2014). Application of the multi-attribute value theory for engaging stakeholders in groundwater protection in the Vosvozis catchment in Greece. *Science of the Total Environment, 470-471*, 26–33. doi:10.1016/j.scitotenv.2013.09.008.

Stillwell, W. G., Von Winterfeldt, D., & John, R. S. (1987). Comparing hierarchical and nonhierarchical weighting methods for eliciting multiattribute value models. *Management Science, 33*, 442–450. doi:10.1287/mnsc.33.4.442.

Straton, A. T., Jackson, S., Marinoni, O., Proctor, W., & Woodward, E. (2011). Exploring and evaluating scenarios for a river catchment in Northern Australia using scenario development, multi-criteria analysis and a deliberative process as a tool for water planning. *Water Resources Management, 25*, 141–164. doi:10.1007/s11269-010-9691-z.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853–870. doi:10.1037/0022-3514.78.5.853.

Van Ittersum, K., Pennings, J. M. E., Wansink, B., & Van Trijp, H. C. M. (2007). The validity of attribute-importance measurement: A review. *Journal of Business Research, 60*, 1177–1190. doi:10.1016/j.jbusres.2007.04.001.

Von Nitzsch, R., & Weber, M. (1993). The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science, 39*, 937–943. doi:10.1287/mnsc.39.8.937.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

Von Winterfeldt, D., & Fasolo, B. (2009). Structuring decision problems: A case study and reflections for practitioners. *European Journal of Operational Research, 199*, 857–866. doi:10.1016/j.ejor.2009.01.063.

Weber, M., & Borcherding, K. (1993). Behavioral influences on weight judgments in multi-attribute decision making. *European Journal of Operational Research*, 1–12.

Weber, M., Eisenführ, F., & Von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34*, 431–445.

Wenstøp, F., & Seip, K. (2001). Legitimacy and quality of multi-criteria environmental policy analysis: A meta-analysis of five MCE studies in Norway. *Journal of Multi-Criteria Decision Analysis, 10*, 53–64. doi:10.1002/mcda.289.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, California: Sage publications, Inc.

Zardari, N. H., Naubi, I. B., Roslan, N. A. B., & Shirazi, S. M. (2014). Multicriteria approach for selecting the most vulnerable watershed for developing a management plan. *Journal of Water and Land Development, 23*, 61–68. doi:10.1515/jwld-2014-0030.

Zheng, J., Egger, C., & Lienert, J. (2016). A scenario-based MCDA framework for wastewater infrastructure planning under uncertainty. *Journal of Environmental Management, 183*, 895–908. doi:10.1016/j.jenvman.2016.09.027.